



HAL
open science

Production gap analysis: an operational approach to yield gap analysis using historical high-resolution yield data sets

C. Leroux, J. Taylor, Bruno Tisseyre

► **To cite this version:**

C. Leroux, J. Taylor, Bruno Tisseyre. Production gap analysis: an operational approach to yield gap analysis using historical high-resolution yield data sets. 12th European Conference on Precision Agriculture, ECPA 2019, Jul 2019, Montpellier, France. pp.75-81, 10.3920/978-90-8686-888-9_8 . hal-02609781

HAL Id: hal-02609781

<https://hal.inrae.fr/hal-02609781>

Submitted on 16 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Production gap analysis – an operational approach to yield gap analysis using historical high-resolution yield data sets.

Leroux, C.^{1,2}, Taylor, J.², and Tisseyre, B.²

¹SMAG, Montpellier 34960, France

²UMR ITAP, Montpellier SupAgro, Irstea, France

cleroux@smag.tech

Abstract

Yield potential has major interests in agronomical modelling such as for the computation of fertilizer recommendations. Most precise estimates of this potential can be generated with crop modelling approaches. However, such potential is often theoretical because it has never been obtained in the field and requires lots of input data to be computed, e.g. soil characteristics, management practices. In this study, a decomposition of this theoretical yield potential into a series of intermediate and more realistic potentials, denoted here production potentials, is proposed. This work comes along with a simple method relying on historical high-resolution yield datasets to compute site-specific production potentials and production gaps, i.e. the differences between actual yield and production potentials. Application of the methodology to a real database of yield mapping has proved of interest to identify areas with similar production potentials but different production gaps. Advantages and concerns regarding the use of such production potentials are discussed at the end of the study.

Keywords: Production gap, production potential, within-field yield data

Introduction

Yield potential is key in agronomical modelling, especially when it comes to computing fertilizer recommendations or to selecting a relevant seeding density (Grassini et al., 2015). This potential is also of particular interest for farmers and advisors because it provides an indication of the possible improvements in yield that can be obtained or the gap that can be filled if adequate management decisions are considered. Within the scientific community, such studies are referred to as yield gap analyses (Van Ittersum et al., 2012, Silva et al., 2016). Estimates of yield gaps are generally obtained via two different approaches. The first approach usually involves crop modelling and does not explicitly require any yield information once the models are calibrated. It can be argued that this approach provides the most reliable estimate of the ‘total’ yield gap as the models account for interactions among weather, soils and management to simulate the highest theoretical yield possible (under perfect production conditions). However, crop modelling requires the specification of many input parameters and current models are rarely, if ever, suitable for application at a site-specific scale. The second approach, in contrast, estimates potential yield, and the associated yield gap, from observed, objective yield measures, i.e. yields that have already been obtained. There have been several methods proposed for deriving these observation-based yield gap including using boundary-functions, upper percentiles of yield distributions and maximum yields derived from one or a combination of experimental station yield, or growers contests (Van Ittersum, 2012).

Most yield gap studies have intended to compute potential yields and resulting yield gaps at medium to very coarse scales, i.e. field, farm, regional or national level. Modelling approaches cannot yet be applied confidently at sub-field levels, in part because the input data, i.e. soil characteristics, weather, management practices have not been available at finer scales and in part because the models are not designed to account for the stochastic noise in these fine-scale measurements. Statistical approaches have utilized average yields at farm/field/regional scales, as these are widely available and have a long temporal sequence. The literature¹ shows that very few studies have aimed to perform site-specific yield gap analysis, despite much of the literature highlighting the potential for precision agriculture approaches to be used in yield gap analyses. This is a clear gap in knowledge as it has been shown multiple times that yield variations are very important at intra-field scales. Nothing prevents the potential yield to be as, or maybe more, spatially variable than the actual yield itself.

The availability of combine yield monitors since the early 1990's means that long time-series (10+ years) of yield data are now available in many arable production systems. This within-field data is very valuable as it provides an objective and site-specific measure of the production that has been obtained across a field. Given the availability nowadays of these temporal site-specific data, a statistical approach to compute and analyze an operational site-specific yield gap, referred to as the production gap in this study, should be possible. This study provides a step towards this goal by deriving a site-specific production gap to help to make more informed management decision at the within-field scale.

Production gap vs. Yield gap

To clarify the notions that will be introduced in this work, the yield gap concept is presented as a response to increasing availability of yield-limiting factors, e.g. water, nutrient, in a two-dimensional plot (Fig. 1). In reality this is a multi-dimensional interaction, but for simplification a two-dimensional representation is given. The Yield gap is usually defined as the difference between the actual yield (A) within the field and the potential yield (F) that the crop might reach under optimal growing conditions, i.e. when water and nutrients are not limited, and when the crop is not subjected to any abiotic or biotic stress (van Ittersum et al., 2012; Grassini et al., 2015). Under those conditions, the crop growth is only controlled by temperature, lightning conditions, crops genetics and atmospheric CO₂. Following these conditions, a theoretical yield response can be derived through crop modelling (Fig. 1, dashed line). This response curve generally forecasts higher yields than those that are actually observed in fields.

When considering empirical, rather than modelling approaches, yield and production gaps need to be determined by pairing and contrasting a target site, i.e. the site associated with A in Fig 1, with sites with similar production conditions. These sites are termed 'relatable' sites and need to be defined within the spatial scale of the analysis. The concept of 'relatable' sites is deliberately scalable in space. For example, they could be applied to i) mean farm yields and bench-marking at a regional scale, ii) mean field yields within a single or grouping of farming systems or iii) site/zone-specific yields at a within-field scale. Be aware that the

¹ A Boolean search of "precision agriculture" AND "yield gap"

following relational yield potentials and gaps described will vary across crops, growers, climatic zones or territories (Figure 1 is a schematic only). To provide more realistic and feasible estimates of yield gaps, some authors have proposed to decompose the yield potential (F) into a series of intermediate potentials (van Dijk et al., 2017; Silva et al., 2016). The theoretical yield response (Fig. 1, dashed line) is complemented by a more realistic boundary yield response (Fig. 1, continuous line) that can be created with ‘local’ yield observations within a given area (van Dijk et al., 2017). Figure 1 is inspired from what was proposed by van Dijk et al. (2017) but additional notions were included, more especially those that pertain to the production gap. All the yield potentials and the associated gaps plotted in Figure 1 are detailed below:

- [A] *Actual Yield* (Y_A): Yield obtained at a site S_i for a given set of inputs Z_A in a given year y_m . It relies on management decision, climatic conditions and crop stresses on yield.
- [B] *Agronomic production potential* (Y_{AP}): For a set of inputs $Z_B (\neq Z_A)$, this is the highest yield that has been obtained in y_m at a site relatable to S_i . Differences with A come from variations in the set of inputs and/or management activities.
- [C] *Climatic production potential* (Y_{CP}): For the level of inputs $Z_C (= Z_A)$, this is the maximum yield that has been obtained at a site relatable to S_i under differing climatic conditions than in y_m .
- [D] *Innovation production potential* (Y_{IP}): For the level of inputs $Z_D (= Z_A)$, this is the maximum yield that has been obtained at a site relatable to S_i with appropriate changes in technologies and management practices. Climatic conditions are the same than for C.

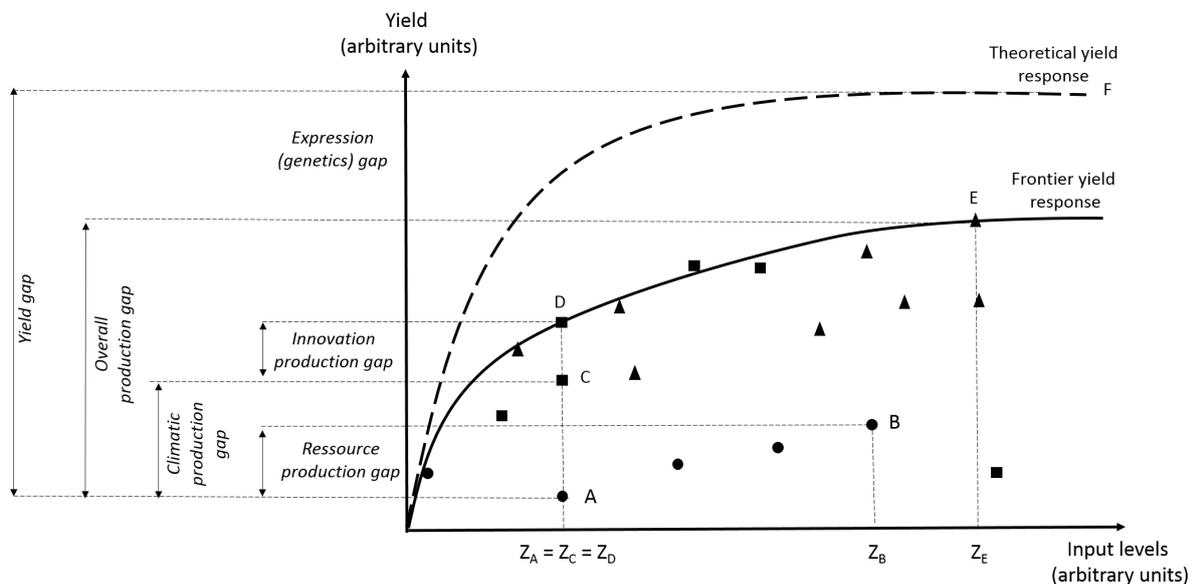


Figure 1. Decomposition of the theoretical potential yield into more realistic potentials. Inspired from van Dijk et al. (2017). Points are indicative of observed yield obtained across space and time. Different shapes indicate different years. **A**: Actual Yield, **B**: Agronomic production potential, **C**: Climatic production potential, **D**: Innovation production potential, **E**: Total production potential, **F**: Yield potential. The Z_i are different sets/levels of inputs.

- [E] *Total production potential* (Y_{TP}): This is the maximum potential yield that has been observed at a site relatable to S_i with an optimal level of inputs Z_E and management (and

favourable climatic conditions). Differences with the potential A come from differences in the set of inputs, management activities, and/or climatic conditions

- [F] *Yield potential* (Y_P): Maximum yield that could be obtained under optimal growing conditions, optimal inputs/management and in the absence of any stress (biotic or abiotic). This potential is theoretical and derived from crop modelling approaches.

Differences or gaps between these potentials can be understood as follows:

- *Resource Production Gap (B-A)*: Potential yield that could have been gained with better input management and management practices. In a SSCM context, this could be derived using data derived from in-field experiments/decisions using variable-rate technologies.
- *Climatic Production Gap (C-A)*: When Z_A tends to be temporally uniform, observed differences in yield at site over time will be due to season-specific weather impacts (e.g. water supply via precipitation) considering relatively uniform management practices
- *Innovation Production Gap (D-C)*: When Z_A tends to be temporally uniform, it presents a plausible potential improvement in production via innovation in technologies or agronomic management (e.g. improve fertilizer type).
- *Overall Production Gap (E-A)*: Potential production gain for a given yield observation relative to all available data and with all options available in regards to improving input levels/efficiencies and improving technologies and management.
- *Expression (Genetic) Gap (F-E)*: Classical difference between (i) potential yield as modelled by the best interactions between crop genetics, weather, soil and management, and (ii) maximum yield observed.

Note that the gap between D and E is not defined as it is a combination of the previously defined production gaps. Here, the focus revolves around the overall production gap, and so for multiple reasons. Firstly, yield data sets provide yield potentials that are realistic and have already been observed, meaning that they could be attained again. Secondly, these data have not previously been well exploited in this context but offer an enormous opportunity for understanding site-specific production gaps/potentials. It is not claimed that the theoretical yield gaps/potentials are not of interest, but rather that the proposed site-specific production potentials require less input data to compute and are collected at and relevant to the scale of analysis (site-specific), which crop models are currently not. However, if the expertise and the relevant data and models are available for sub-field modelling, theoretical potentials might be equally valuable or preferable to calculate. In this case study, information on the amount of inputs applied is incomplete, thus it must be assumed that the set of inputs Z is not uniform over seasons. If it was, then the analysis would only be of the Climatic and Innovation production gaps.

Material and methods

Description of the multi-temporal within-field yield data

The study was carried out on a 31-ha field in the north of France near Evreux (WGS84 datum: E: 0.78, N: 48.95). The field is cropped in a wheat (2003-2005-2007-2009-2011-2013-2015) and canola (2004-2006-2010-2014) - flax (2008 - 2012) rotation, i.e. a ‘first’ wheat is grown

every other year with a break crop in between. Nine years of yield mapping were available spanning the 2003-2015 period but only the wheat crop was considered in this study (six years) to minimize the need to standardize between different crop types in this exploratory work. More descriptive, spatial and temporal characteristics can be found in Leroux et al. (2018).

Yield map pre-processing

Within-field yield data were pre-processed in a similar way as in Leroux et al. (2018). Outlying observations were first filtered, and a grid composed of 10x10m pixels, whose orientation followed that of the harvested rows, was then superimposed on the yield data. For each pixel of the grid, yield values were first averaged by year so as to obtain one yield value for each pixel and each year. The objective was to make sure that each year had the same influence in each pixel even if the number of observations falling into each pixel was different from year to year. Empty pixels in specific years due to missing yield observations were given the mean yield value over the years in the same pixel.

Computation and evaluation of the yield overall production gap

The estimated yield total production potential (\widehat{Y}_{TP}) is defined over each pixel i and is computed as the maximum yield obtained on the i^{th} pixel throughout the years:

$$\widehat{Y}_{TP}(i) = \max(Y^m(i)), \quad m \in \{1: M\} \quad \text{Eq. 1}$$

Where $Y^m(i)$ is the i^{th} yield observation in year m , and M is the number of years in the historical yield database².

The year at which this potential is reached is denoted m' in the rest of the study. Once this potential is defined, the estimated annual overall production gap (\widehat{Y}_{pg}^m) can be easily computed over each pixel i for each year m as the difference between the total production potential and the actual yield (Eq. 2).

$$\widehat{Y}_{pg}^m(i) = \widehat{Y}_{TP}(i) - Y^m(i), \quad m \in \{1: M\} \quad \text{Eq. 2}$$

Where $\widehat{Y}_{pg}^m(i)$ is the overall production gap of the i^{th} yield observation in year m , and M is the number of years in the historical yield database.

The annual overall production gap is interesting in itself as it can be used to evaluate the impact of specific growing conditions or management practices during a given year.

² In this case, the ‘reliable’ sites are simply constrained to the same pixel over time. However, there is no reason why Y^m could not be derived from a vector of yield values from ‘reliable sites’ in year m , instead of just the local pixel value (if sites with similar production conditions were known and the associated data was available).

However, to fill such gap, there is a need to answer operational questions such as (i) how often this potential is reached or almost reached? or (ii) how much deviation from that maximum is there in a given pixel over time? In this study, the second option is chosen and a multi-annual overall production gap \widehat{Y}_{pg} is proposed to be computed over each pixel i using Eq. 3 as:

$$\widehat{Y}_{pg}(i) = \frac{1}{\widehat{Y}_{TP}(i)} \times \frac{1}{M-1} \times \sum_{m=1, m \neq m'}^M \widehat{Y}_{pg}^m(i) \quad \text{Eq. 3}$$

This indicator is expressed as a percentage so that it can be compared across varying situations. It is computed over $M - 1$ years because the year m' is not taken into account.

Results and discussion

Characterization of the within-field production gap over the field under study

Figure 1 details the annual spatial wheat yield patterns (2003-2015) on common scale for all years. It is clear that some years are much more productive, e.g. 2009 and 2015, than others. Some yield patterns also exhibit a much wider range of variation, e.g. 2011, than others that seem to be much more homogeneous, e.g. 2005 and 2007. Spatial yield patterns are also relatively stable from year to year, especially in 2003, 2005, 2009, 2011 and 2015, with a lower productive area on the northern section of the field and a higher-yielding one on the centre and southern portions of the field (Leroux et al., 2018).

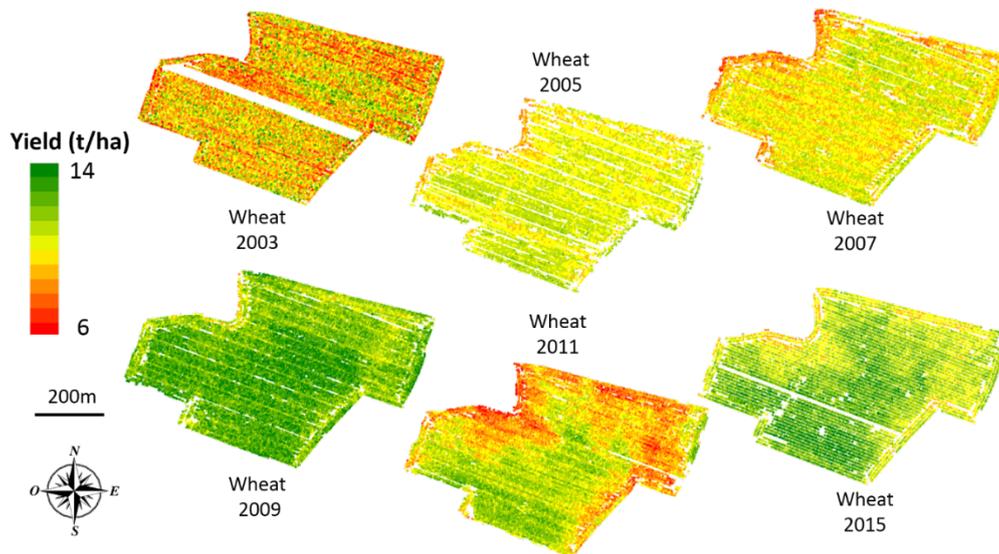


Figure 2. Within-field wheat yield maps of the field across the 2003-2015 period.

The yield total production potential and multi-annual overall production gap were computed from this historical database of yield mapping. Unsurprisingly, the production potential spatial pattern matches the yield spatial patterns in Fig. 2. This operational potential is relatively variable over the field and ranges from less than 10 to more than 13 t/ha. This information

should drive boundary decisions on maximum and optimum input levels and/or investigations into yield determining factors at each site. From a more general perspective, the total yield production potential in this field is strongly related to the yield obtained in the most productive year (2009), although it is unclear if a single year dominance like this is likely in other fields and production systems. The multi-annual overall production gap, on the contrary, exhibits a different spatial pattern that is not linked with the previously described yield spatial patterns (Fig. 3). Areas of both high and low yield potential exhibit a low gap in production. There is however a central zone in the field where the production gap is relatively high (> 15%) that incorporates both high and low yield potential areas. The multi-annual overall production gap brings in operational information to enables users to identify and manage areas that have the highest potential for improvement (or conversely, the highest area of risk). For example, it is clear that the northern section of the field is the less productive, but also the less variable. If the yield-limiting factor here is not manageable, it may mean that it is more valuable to put efforts into reducing the production gap in the centre of the field.

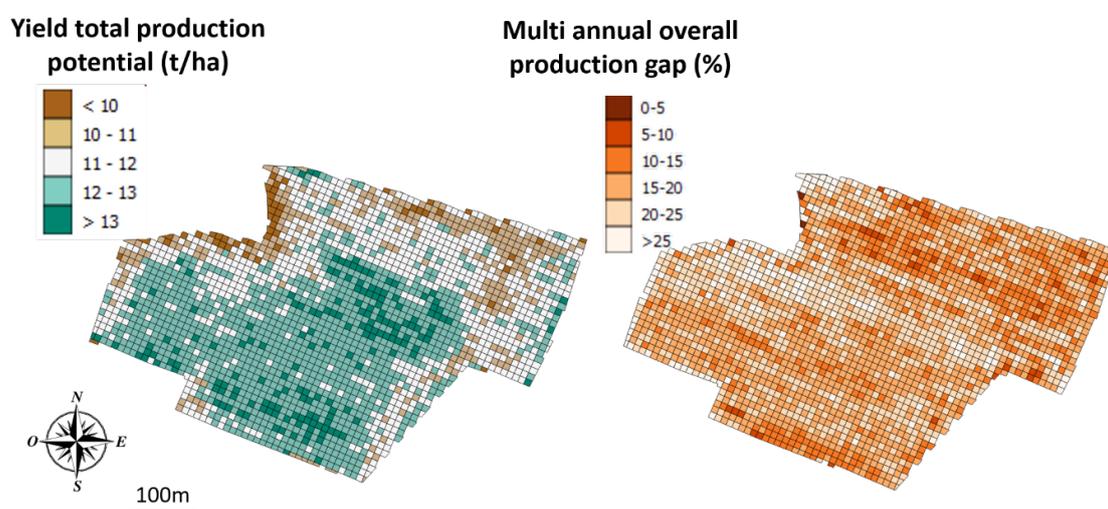


Figure 3. Within-field yield maps of the yield local production potential and multi-annual local production gap across the 2003-2015 period.

Further considerations

It must be stressed that the proposed production gap analysis is iterative in nature as many factors are likely to evolve across time, e.g. technologies, climate, variety. The intent of the production gap analysis is not to assess whether a technological change should be done but rather to say that with such technology or such variety, a given yield can be obtained. It remains possible (would be expected at some stage) that the actual yield of the current season outreaches the historical production potential. It is suggested here that the yield local production potential would simply be replaced by the yield of the current season, which would be seen as the new local production potential. In such case, it should not be considered that the previously proposed production potential was a bad estimate but rather that as more years of yield mapping become available (best available data), estimates become more viable as they encompass a large range of growing conditions and can be derived from years with climatic conditions similar to current (or forecasted) conditions. It is nonetheless

acknowledged that the statement of best available data is questionable as crop models do exist and provide higher and certainly more accurate yield potentials. However, the strong advantage of historical yield maps is that they provide an operational potential, i.e. yield values that have already been obtained.

The (site-specific) yield total production potential could be seen as a yield objective at the beginning of the cropping season. This potential should be revised during the growing season as uncontrolled environmental factors evolve, e.g. weather effects or disease pressures. One issue, however, with the proposed production gap analysis is that growers might obtain each year a yield that is close to their local maximum yield, but without knowing if their management practices are optimal. In fact, as this local maximum is seen as a production potential, growers cannot know if they can possibly outreach it. This information should be used to design and perform site-specific experiments to gain a better insight into the local yield response curve to inputs and management practices. In this study, a specific attention was paid to the evaluation of the within-field yield total production potential, mainly because information on the amount of inputs applied or precise management practices is incomplete. Production potentials are interesting because they can be computed in terms of gross margins and gross margin gaps to help users evaluate the economic risk of making future management decisions. It must be clear that different yield production potentials would necessarily lead to different gross margin-related maps (Fig. 1).

Conclusion

In this work, an operational approach relying on historical databases of yield mapping was proposed to help farmers position themselves in terms of their actual yield levels and their potential for improvement. The objective being to deliver maps of total production potential and multi-annual overall production gap. The proposed approach is iterative in nature as new yield maps will improve the definition of the yield production potentials and gaps. Future works will aim at merging the production gap analysis with economics to evaluate the risk for farmers of making future management decisions.

References

- Grassini, P., van Bussel, L.G.J., van Wart, J., Wolf, J., Claessens, L., Yang, H. et al. (2015). How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. *Field Crops Research*, 177, 49-63.
- Leroux, C., Jones, H., Taylor, J., Clenet, A., Tisseyre, B. (2018). A zone-based approach for processing and interpreting variability in multitemporal yield data sets. *Computers and Electronics in Agriculture*, 148, 299-308
- Silva, J.V., Reidsma, P., Laborte, A., & van Ittersum, M.K., (2016). Explaining rice yield gaps in Central Luzon, Philippines: an application of stochastic frontier analysis and crop modelling. *European Journal of Agronomy*
- Van Dijk, M., Morley, T., Jongeneel, R., van Ittersum, M., Reidsma, P., & Ruben, R. (2017). Disentangling agronomic and economic yield gaps: An integrated framework and application. *Agricultural Systems*, 154, 90-99.

Van Ittersum, M.K., Cassman, K.G., Grassini, P., Wolf, J., Tittonell, P., & Hochman, Z. (2012). Yield gap analysis with local to global relevance—A review. *Field Crops Research*, 143, 4-17.