Aldo Benini
A note for ACAPS

The use of Data Envelopment Analysis

to calculate priority scores in needs assessments

8 July 2015

Contents

Tables and figures in the main body of the note

Summary

What this is about

In humanitarian action, "*setting priorities is part of strategic response planning*" (UNOCHA 2014). This is a difficult task. Numerous tools have been developed to aid it, but a universally suitable algorithm to establish priority indices has not been established.

This note introduces a method known as Data Envelopment Analysis (DEA) as a tool for computing priority measures for affected communities, sites or social groups. It is attractive because it circumvents some of the issues in classic index formation altogether. It avoids pre-defined indicator normalization, weights and aggregation functions in favor of data-driven parameters. It is attractive also because a freeware application is available that works closely with MS Excel, the workhorse of humanitarian data analysis.

DEA gives the "benefit of the doubt"

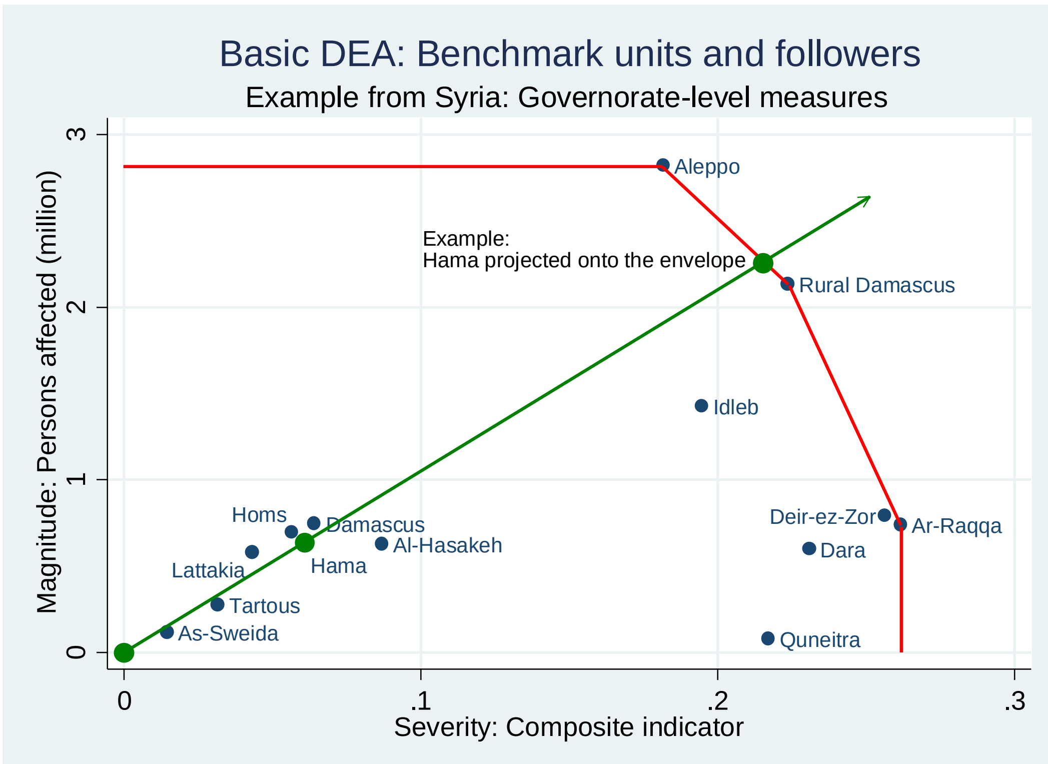
DEA is suitable particularly for relatively simple situations - situations in which only two or three measures (primary or composite) are to be combined. In the simplest case of two measures - say, the number of affected persons for magnitude, and their proportion in the pre-crisis population for intensity - DEA offers an immediate and intuitive visual interpretation.

Known also as the *"benefit of the doubt method"* in social indicator research, DEA translates a simple philosophical assumption. The name reflects the intuition that a unit (area, social group, individual) may attain a high score on the strength of a *high value on one indicator even if it is low on some or all of the others*. This one high value may justify enough of a "benefit of the doubt" in order to assign a high score to this unit.

DEA side-steps the "apples and oranges" problem

Technically, DEA proceeds by selecting a set of benchmark units which, by construction, all receive the highest score. The scores of their followers depend on their relative positions vis-à-vis the benchmark units. In other words, DEA produces relative measures. This is a strength as well as a weakness. DEA side-steps two classic problems (normalization and weighting) at the expense of inviting new ones of robustness and

aggregation. It is not robust to outliers; results depend on the choice of administrative levels. For a first visual impression, this graph illustrates the basic elements – benchmark units, which together define the set of indicator combinations with the highest scores (the red envelope); followers defined by the positions vis-à-vis the envelope (points inside).



The graph is repeated and discussed in detail in the main body of the note.

DEA obviates the need for indicator weights

As a result, we believe that DEA occupies a valuable, if narrowly circumscribed niche in prioritization methods. Compared to additive and multiplicative index formulations, DEA is at its best when analysts wish to hold off on *apriori* weights or aggregation functions. This situation obtains frequently in the early stages of crisis response – if and when analysts have the humility to admit their scant understanding of what determines severity. As time passes, and the crisis impacts and dynamics are better understood, the case for classic indices grows, and the comparative benefits of DEA may fade. In either method, analysts ought to work out a clear process model of how pre-existing conditions, magnitude and intensity cooperate in producing the observed crisis impacts. The model, and thus the prioritization formula, should be driven by the plausible causal nexus, and not chiefly by data availability.

An invitation for busy readers

The note proceeds through several sections, with the double objective of giving the reader a firm conceptual grounding (without the DEA mathematics) as well as a step-by-step walk through the freeware application OS-DEA. Specifically, we elaborate the rationale for DEA, defining situations where this method is appropriate, and others in which classic (additive, multiplicative) index formulations are preferred. Two case studies follow. The first demonstrates a DEA application with a small number of units – the 14 governorates in Syria - in an intuitive geometric interpretation. The second develops a more ambitious model with informative variables on both the input and the output side. The data are from over 400 communities in the Philippines struck by a typhoon. This section emphasizes the right choice of the DEA model as a function of the process model connecting disaster and humanitarian impact. The distributions of priority scores using DEA vs. a multiplicative aggregation are compared. We then use the same data to demonstrate the DEA computations with a free application, OS-DEA. The reader wishing to replicate them finds the data and our results in the demo workbook.

Finally, we discuss three general issues that DEA users will frequently face. One of them – the treatment of pre-existing conditions in priority scoring – is generic, regardless of DEA or other index methods. Two – the treatment of cases with genuine zeros in the indicator values, and the discrepancy between DEA results depending on the administrative level of the original measurement – are specific to it.

We conclude by inviting the reader to make a modest learning investment in DEA and to practice a comparative attitude in which intelligent questions arise when we compare results obtained with DEA to those of traditional methods. It will take several small experiments, followed by at least one major needs assessment using DEA prominently and successfully, before this method can claim parity with the established methods. A growing community of users, or at least sympathizers, will accelerate the process.

Introduction

For many crisis-affected populations, humanitarian aid is in short supply. Yet even when the response is abundant, matching resources with the most urgent needs is not always straightforward. The task of humanitarian information management is to connect needs assessment and resource allocation. In recent years, the community of practitioners has experimented with a number of prioritization tools, within an evolving common methodology, but as yet without a universally suitable algorithm to establish priority indices.

Common practice involves the construction of composite measures - priority or severity indices - from indicator data at hand, the classification of units (areas, sites, groups) by index levels and representations of the affected regions in graphs and maps that show those levels. These representations then figure prominently in reports and in humanitarian update and dashboard products. (In this note, we will use priority and severity interchangeably).

The indices have some commonalities, but they also vary widely in substance and mechanics. They commonly incorporate measures of vulnerability (of the population) as well as of magnitude and intensity (of the disaster or crisis). They vary in several aspects. Some may address only one or two of the three key concepts vulnerability, magnitude and intensity. This happens when good measures for one or the other are not available, or when primary indicators are combined in ways reflecting other concepts. Formally, the indices differ in the mathematical operations that lead to composite measures - the normalization and weighting of indicators, and their aggregation.

Observers of these practices have raised a number of criticisms. Some were first made outside the humanitarian field while others have been motivated by the observed implementation of the severity or priority indices. Development researchers object to "mash-up indices" that are primarily driven by the availability of data, to the neglect of a solid grounding of what the data and concepts mean (Ravallion 2010). For humanitarians, this can be particularly delicate when their indices incorporate the lives lost in the crisis, as some do. Mixing mortality with other indicators implies a price on human life, through the substitution rates that their weights define. "Ten destroyed buildings have the same effect for priority as one life lost does", or anything distantly resembling this, is hardly the kind of trade-off with which an agency would like to be publicly identified.

A second criticism concerns transparency and statistically defensible operations once the indicators have been selected. Indices have differed greatly in how they wired these variables together, what intermediate steps were taken, and which of the transformations were statistically legitimate and, if unusual, came with explicit rationales. Egregious abuse has been committed particularly of ratings and rankings. Rarely has the robustness of the resulting priority scores been tested against changes in indicator weights, intermediate transformations or measurement error. However, it is fairly common to see differences discussed between population-weighted and unweighted priority or severity scores. Benini (2012) discusses those and related issues in general; Benini and Chataigner (2014) dissect them in four needs assessments in the Philippines, proposing improvements and alternatives for future needs assessments anywhere.

This note wishes to raise awareness about a method that produces severity or priority scores while circumventing some of the issues in classic index formation altogether. "Data Envelopment Analysis" (DEA) has transpired into social indicator research from an unexpected origin, the study of economic efficiency. Used with social indicators (such as to measure social deprivation), DEA is known as the "benefit of the doubt" method. It avoids pre-defined indicator normalization, weights and aggregation functions in favor of data-driven parameters. In fact, it generates weights that differ, not only between indicators, but also among the units (affected groups, geographical areas, etc.).

Several considerations motivate this note. One of them is the drive in the humanitarian community to *"generat[e] structured information products that can facilitate joint intersectoral analysis of humanitarian needs"*. This is evident in, among other initiatives, the "Humanitarian Needs Comparison Tool", developed by UNOCHA for the 2015 Humanitarian Needs Overview (UNOCHA 2014). The endeavor is likely to enhance receptivity for diverse analytic approaches.

Second, as often in research, new methods become practical once new software is available on familiar platforms. Some DEA applications are free, and recently one has appeared that works closely with, if not yet in, Excel, the workhorse of humanitarian information managers. This agrees with a third consideration, which is speed and convenience, particularly in the early, rapid stages of needs assessments.

Above all else, however, we recommend DEA in situations where needs assessment analysts wish to avoid, or at least postpone, difficult choices of pre-set indicator weights and aggregation functions. DEA is suitable particularly for relatively simple situations - situations in which only two or three measures (primary or composite) are to be combined. In the simplest case of two measures - say, the number of affected persons for magnitude, and their proportion in the pre-crisis population for intensity - DEA offers an immediate and intuitive visual interpretation.

DEA is not recommended when analysts desire to establish data-driven weights for a larger number of indicators (there are other methods for this). The robustness of its results to measurement error or sampling variance is visually accessible in the case of two measures only. For higher dimensionality robustness tests require simulation methods that are not straightforward. Other limitations will be discussed further below.

The rest of this note proceeds as follows: The next section elaborates the rationale for DEA, defining situations where this method is appropriate, and others in which classic (additive, multiplicative) index formulations are preferred. Two case studies follow. The first demonstrates a DEA application with a small number of units – the 14 governorates in Syria - in an intuitive geometric interpretation. The second develops a more ambitious model with informative variables on both the input and the output side. The data are from over 400 communities in the Philippines struck by a typhoon. This section emphasizes the right choice of the DEA model as a function of the process model connecting disaster and humanitarian impact. The distributions of priority scores using DEA vs. a multiplicative aggregation are compared. We then use the same data to demonstrate the DEA computations with a free application, OS-DEA. The reader wishing to replicate them finds the data and our results in the demo workbook[1]. Finally, we discuss three general issues that DEA users will frequently face. We conclude by repeating that DEA is at its best when analysts wish to hold off on *apriori* weights or aggregation functions. DEA-based priority scores may be an intermediate construct, to give way to more classic indices once the analyst has a firmer understanding of the crisis environment and the causal nexus determining the observed and measured humanitarian impacts.

The rationale for DEA

Criteria for good methodologies

If we accept that the criticisms of classic index formation are valid, we need to re-direct the search for good humanitarian priority measures. We want indices that are relatively simple, transparent, have good conceptual validity, require a minimum of arbitrary choices, and are likely to meet with the consensus of practitioners:

- "Simple" means that non-statistician information officers can do the work, notably in Excel or at most with the help of a quick-to-learn Excel add-in.
- "Transparent" means that users, with a minimum of guidance, can understand the mathematical concepts and operations, either in formulas and functions, or in helpful visual models.
- "Conceptual validity" means that constructs are built on some theoretical tradition that has proven productive, if not yet in humanitarian data management, at least in other disciplinary fields.
- "Minimizing arbitrary choices" means that in the triathlon of normalization, weighting and aggregation better informed choices are made than in "classic"

[1] Currently it consists of four files within “Acaps_DEA_DemoFiles_150707.zip”.

index formation. This can happen by making specific choices on the strength of some relevant theory or experience, or by side-stepping some choices altogether.

- "Likely to meet the consensus of practitioners" means that inputs, mechanics and results of the priority measure appeal to members of this community, if not in every detail, then at least in the larger intuition.

At the conceptual level, those desirable properties have been key to recent developments in the theory of social indicators, under headings like "benefit-of-the-doubt method" (Cherchye, Moesen et al. 2007) or "let the data speak for themselves". Technically, they are translated by the DEA algorithm.

We will briefly characterize the "benefit -of-the-doubt" philosophy, then help the reader determine situations suitable for DEA with the help of decision diagrams, and later present a minimum of DEA technicalities in a visual approach.

The "Benefit-of-the-Doubt" method

The "benefit-of-the-doubt" method means the application of a particular statistical model to problems of composite measures. The values - the "index scores" - are computed by a variant of Data Envelopment Analysis (DEA) (see below).

Behind it is a philosophical assumption. The name reflects the intuition that a unit (area, social group, individual) may attain a high score on the strength of a *high value on one indicator even if it is low on some or all of the others*. This one high value may justify enough of a "benefit of the doubt" in order to assign a high score to this unit. Such methods are easily available outside DEA, and in fact are part of everyday life. For example, if the indicators are commensurate, one can simply compare positions with respect to maxima and minima. A person shopping for soft drinks may consider unit price ($ per litre, minimize) and sugar content (also minimize) and hence arrive at a priority order reflecting best subjective compromises. DEA does not need prior commensuration and treats deviations from the extremes in formalized ways.

Substantively, in the humanitarian domain, one may argue that the severity of a local situation is a function of magnitude and intensity. A high priority is warranted if the affected area or social group is either large or critically needy or both. In addition, vulnerability is a consideration that, where possible, should be modeled too.

The rationale is that even in large units with, on average, a less than critical level of need, the intensity of need internally varies, and some persons will likely be critically needy. Conversely, in small units with on average critical needs, critically needy persons may be as many as, or more than, such persons in large units with overall lesser needs. Thus, although we may be unable to determine specific weights for any magnitude and intensity indicators, the intuition is that magnitude and intensity may compensate for each other in justifying humanitarian priorities. How this compensation works in DEA will be discussed further below.

Origins and suitability for humanitarian analysis

DEA was pioneered in the 1970s as a new approach to measuring the efficiency of programs and organizations (Cooper, Seiford et al. 2007)[2]. It was developed to define and compute efficiency scores for production processes in which some or all of the inputs and output could not be measured in monetary terms. As such, it confronted a typical apples-and-oranges problem. The challenge that it faced was to make those variables commensurate without imposing particular normalizations or fixed weights. DEA models overcame it by defining a set of benchmark units (top performers) and, for the remainder of the units, a metric of efficiency *relative* to the benchmark. Essentially, therefore, DEA produces *relative* measures (in contrast to absolute financial measures such as profitability); the set of benchmark units and the scores of their followers depend on the ensemble of units and of their measured variables. This is one of its strengths as well as weaknesses; it side-steps two classic problems at the expense of inviting new ones of robustness and aggregation.

DEA has proliferated into a large subfield of economic analysis, with a bevy of specialized models and software applications, of both the high-end commercial and the, by now, comfortably manageable freeware type. Google Scholar, as of February 2015, referenced 56,000 scholarly works on DEA. Despite its early application to a social welfare question - the evaluation of schools with varying poverty and disability burdens - it took some time for researchers to realize DEA's potential for problems of composite measures beyond the productive efficiency perspective. Cherchye et al. (2007, op.cit.) is a seminal article that makes the case for social indicators research, although his illustrations remain tied to the productivity concept[3]. Rather, it is DEA applications in deprivation research, particularly in health care, that build the bridge to humanitarian concerns.

The "trick", if you will, is to consider measures of deprivation as outputs while feeding the software with a symbolic uniform input (conveniently = 1). In this "pure output setting", the efficiency score can take any new meaning, as suits the application - it may be a "social deprivation score" or a "humanitarian priority score" or yet a score with some other substantive meaning.

---

[2] The Wikipedia article on DEA is not particularly good.

[3] For critical remarks on the use of DEA in social indices in general, see Decancq and Lugo (2012). In the humanitarian field, DEA as a technique is primarily used for the optimization of logistics. This usage is closer to the original efficiency concept. See, for an overview, Zinnert, S., H. Abidi, et al. (2011), and for a practical application, Alsharif, K., E. H. Feroz, et al. (2008).

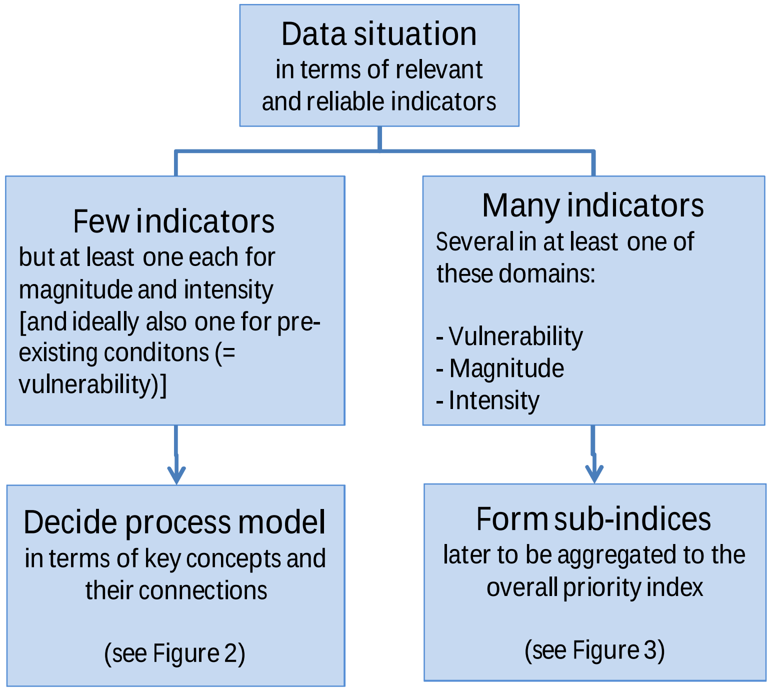
As mentioned, much of the attractiveness of the DEA model rests in the ease with which the distribution of units (areas, social groups) can be visualized as long as we deal with just two outputs - say, one of them a magnitude measure, the other an intensity measure. How each of these measures is formed is a separate question, one which has to be solved before building and estimating the DEA model.

Situations suitable for DEA and for other methods

Three diagrams in this section will stimulate the reader to think about situations for which DEA may likely be appropriate, and others where it does not normally belong. Important for this line of thought is the distinction between *process model* and *measurement model*, further elaborated in Benini and Chataigner (2014). The process model defines how the conceptual elements are connected. In this context it defines how vulnerability, magnitude and intensity interact for greater or lesser impact, and hence severity and priority. The measurement model directs how each of these concepts is to receive a numeric measure using the indicators or other data at hand (or yet to be collected). In the actual work, process and measurement models are conflated by practical necessity and opportunity. Still it is helpful to keep them distinct conceptually, in the interest of pursuing validity and reliability. Both process and measurement models must be valid in order to be useful; reliability is a measurement concern so much so that a good process model can (and should) survive amid inaccurate data, in hopes that subsequent measurements will improve.

The decision path starts opportunistically with the data situation, and with the use by the analyst of only a bare minimum of indicators, or of a larger number that first will have to be grouped and transformed into intermediate quantities (Figure 1).

Figure 1: Basic data situation

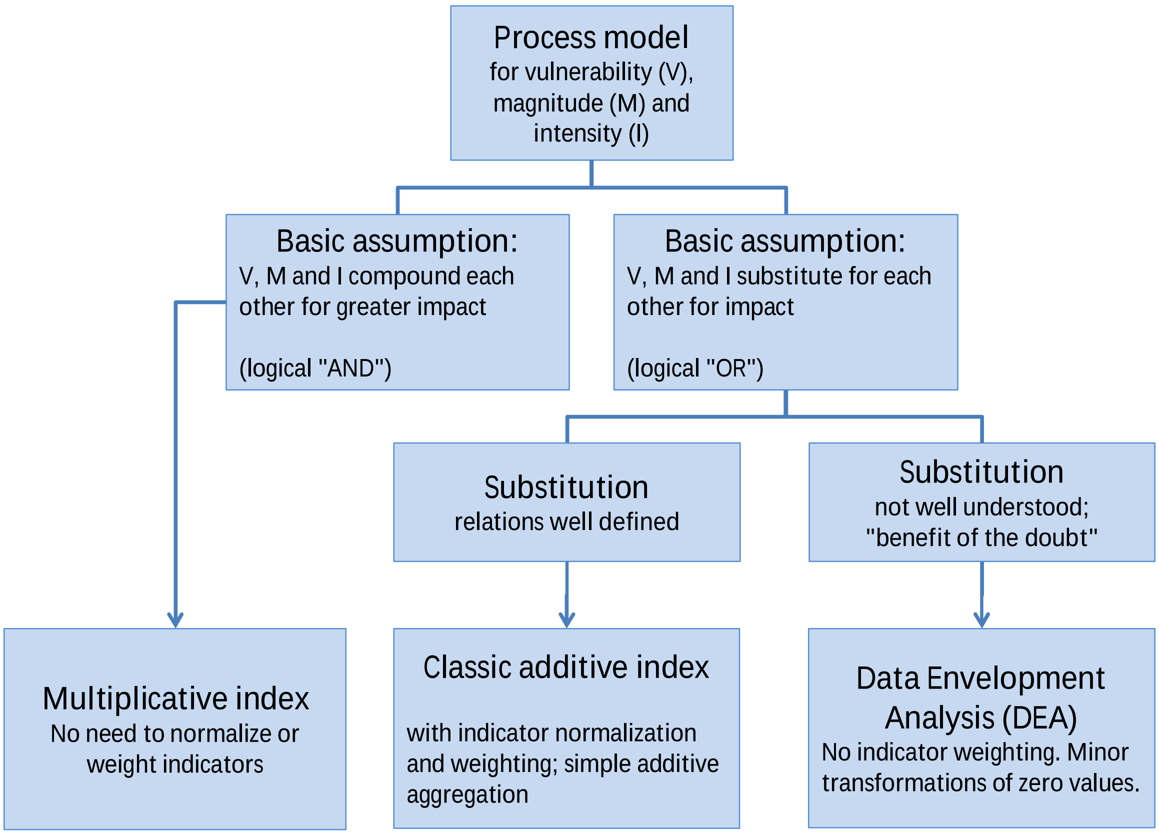


Basic to the process model is the assumption of how vulnerability, magnitude and intensity interact to determine severity and hence priority. A second, similarly important point is the confidence that the analyst places in the indicator weights. The weights, in linear and additive models, determine the substitution rates between indicators for their effects on the severity[4]. DEA is indicated primarily when the analyst does assume substitution, but does not think she understands the processes well enough to warrant constant rates of substitution in any pair of indicators.

---

[4] The point that weights in such models do not express the absolute importance of variables, but determine substitution rates, has often been made, e.g. Nardo et.al. (2008: 112) in the well-known OECD Handbook on Composite Indicators.

Figure 2: The case for DEA



The diagram defines a niche for DEA, given certain assumptions for the process model. In the DEA box, “minor transformations of zero values” is a technical note to be revisited below. It refers to the need to avoid division by zero by replacing observed zeros with small positive values. This is the only transformation needed, and only when there are zero values.

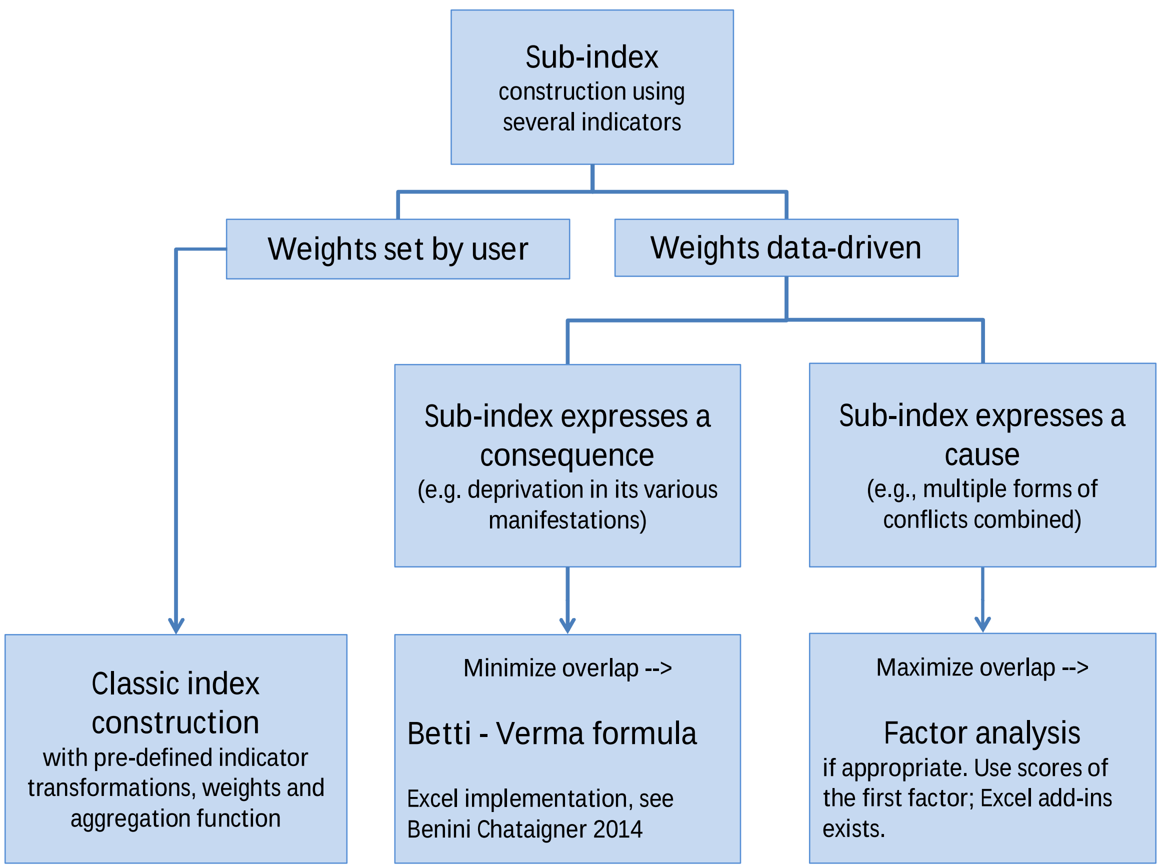
Readers may object that the multiplicative index makes similar substitution claims as the additive index does. The only difference is that the rate is constant in the logarithms of the measures, rather than in the absolute values. This objection is granted. It entails that the crucial decision is between an analyst who is certain how the key processes work (and hence opts for a multiplicative or additive index) and another who is vague about the specifics of substitution (and hence prefers DEA). In this sense, DEA expresses the humility of not knowing how exactly vulnerability, magnitude and intensity interact.

The strategic choices open in the construction of sub-indices have little to do with DEA. Although we have not found this discussed in the literature, we do not foresee situations in which DEA would be preferred to the methods recommended in Figure 3.

Philosophically, doubts about the appropriate effect of an indicator should result in its elimination from the sub-index model rather than in flexible weights. Technically, many primary indicators may have cases with zeros; these would not work in the DEA algorithm unless suitably transformed. We simply note in what situations sub-index construction may be the way to go; if indicated, its technicalities have nothing to do with DEA.

The major choices concern who is to set the indicator weights, and whether one wishes to maximize diversity (i.e., minimize overlap) among indicators, or rather minimize diversity (maximize overlap). The latter choice depends on whether the construct is meant to capture a *consequence* of the indicators, or an underlying common *cause* that expresses itself in the indicators. The Betti-Verma formula recommended for minimizing overlap is discussed in Benini and Chataigner (2014). An Excel implementation is given in the accompanying demonstration workbook and, for convenience, copied in our demo file.

Figure 3: Strategic choices in sub-index construction



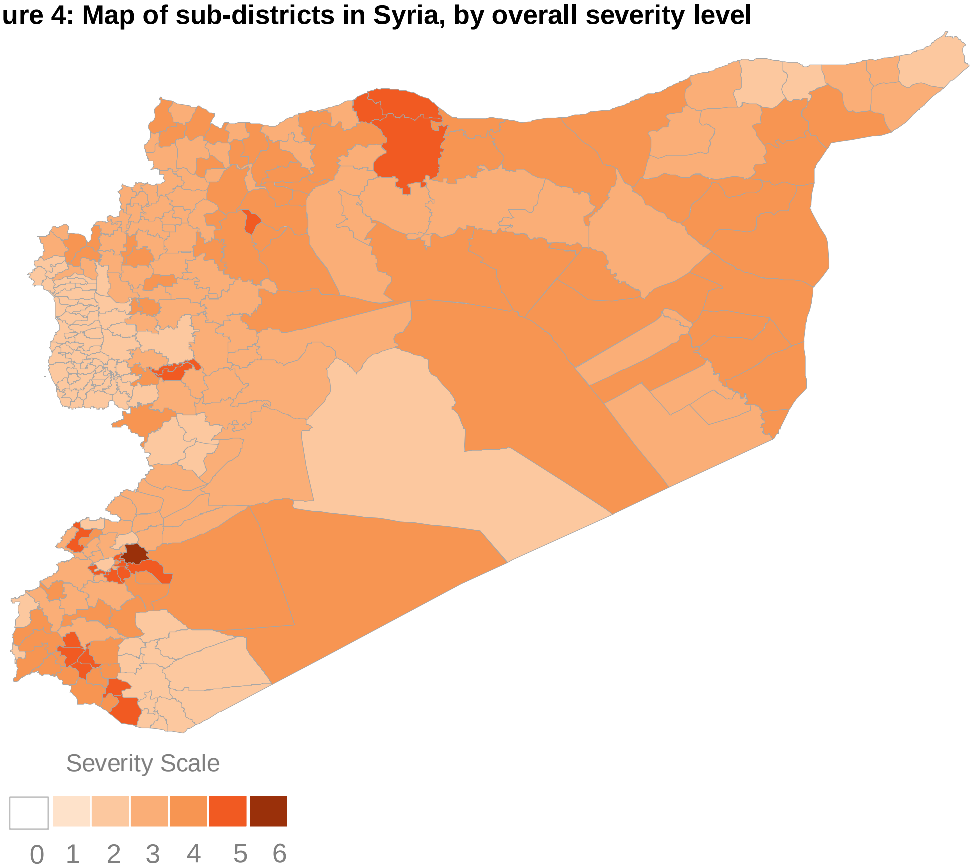
However, *after* all the sub-indices have been computed, nothing stops the analyst from considering DEA as the method integrating them in the overarching process model.

Case study #1: Prioritizing governorates in Syria

Background

The Joint Data Review in preparation of the 2015 Syria Strategic Response Plan for Syria reported estimates of persons in need as well as of IDPs for 270 sub-districts, alongside the latest (2011) pre-conflict population estimates[5]. In addition, the sub-districts were given severity scores in each of the education, food security, health, shelter, and WASH sectors. Severity scores were assigned also for early recovery, humanitarian access, and protection. An overall score was created as the geometric mean of those eight severity scores. The map, from the Humanitarian Needs Comparison Tool (UNOCHA 2014), depicts the sub-districts with their overall scores.

Figure 4: Map of sub-districts in Syria, by overall severity level



Source: Syrian Arab Republic: Humanitarian Needs Comparison Tool, Draft Version, updated 3 Nov 2014.

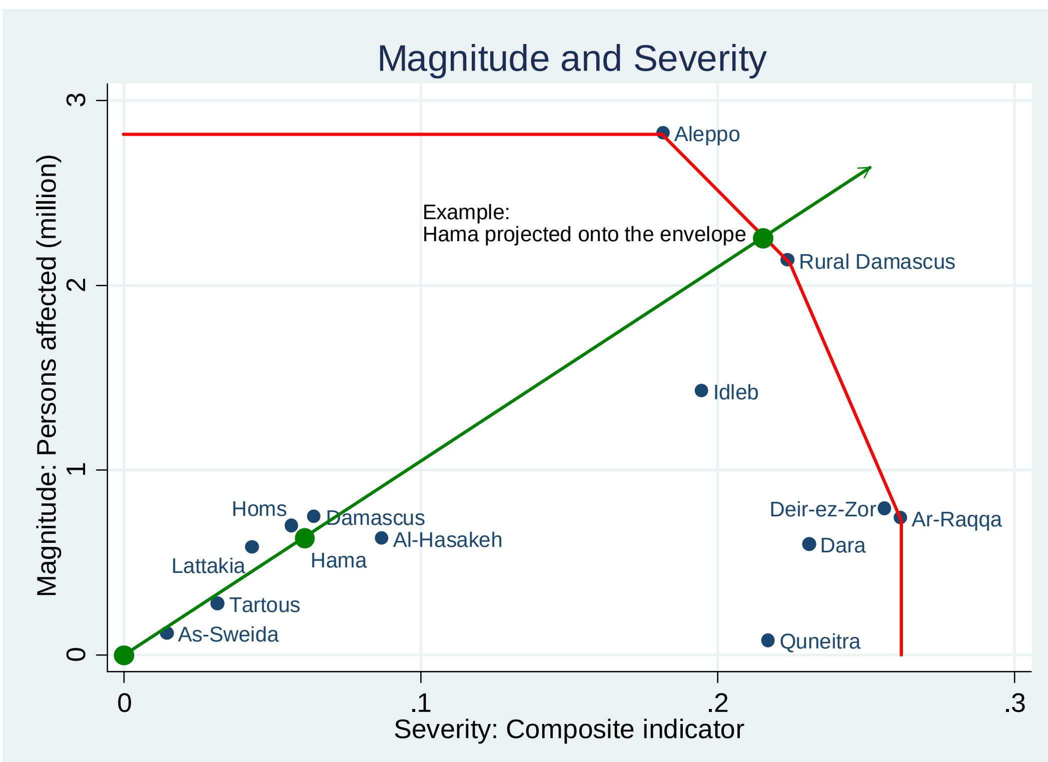
[5] The Joint Data Review database has 272 sub-district records. However, two records, both in Quneitra Governorate, had zero population in 2011 and zero PiN in 2014. We exclude them from further analyses.

The definition of the sectoral scales is not entirely clear. Regardless, by virtue of its fine resolution, down to sub-districts, the 2015 Joint Data Review database is a data mine for priority score modelling. For purposes of the "Regional Analysis of the Syria Conflict (RAS)", 1st October - 31st December 2014 issue (SNAP 2015), we reformulated the magnitude and severity measures. Exceptionally in this model, we use the term “severity” for “intensity” because both the Joint Data Review and the RAS named their intensity measure “severity”. Magnitude was expressed as the number of persons affected, for which the greater of the numbers of IDPs and persons in need (PiN) was taken. The technical details of the severity measure are not of interest for this exposition; interested readers may consult the RAS document. For easier visualization, we report a governorate-level model - there are only 14 such units.

Basic DEA logic

Figure 2 is a scatterplot of the 14 governorates by those two measures - magnitude and severity. The red and green lines are explained below.

Figure 5: The 14 governorates, by magnitude and severity



In this simple scenario, the DEA proceeds as follows:

1. A subset of the governorates is determined, such that
   1.1. For any given member of the subset, none of the other 13 governorates exceeds it on both measures.
   1.2. An envelope (red line) connects the outward governorates and the axis, such that
      1.2.1. it nowhere bends inward (= it is non-convex),
      1.2.2. from the governorate with the maximal value on the severity measure, a perpendicular line drops to the x-axis. Analogously for magnitude and y-axis.

Rule 1.1. qualifies Aleppo, Rural Damascus, Deir-ez-Zor and Ar-Raqqa (Idleb is exceeded by Rural Damascus, Dara by Ar-Raqqa). Rule 1.2.1. excludes Deir-ez-Zor. Rule 1.2.2. creates the horizontal segment from Aleppo to the y-axis as well as the vertical line from Ar-Raqqa to the x-axis. This is relevant for the next steps.

The envelope (hence Data Envelopment Analysis) is sometimes known as the frontier, reminiscent of the "production frontier" in the original efficiency understanding of DEA. In our case, three governorates define the envelope:

- Aleppo on account of its maximal magnitude,
- Ar-Raqqa on account of its maximal severity, and
- Rural Damascus as a mixture of those two considerations.

All three receive the highest priority score, which, by DEA convention, is = 1. Note that Rural Damascus could be anywhere as long as: a. its magnitude is lower than Aleppo's, b. its severity is lower than Ar-Raqqa's, and c. it is on or to the right side of an imaginary straight line connecting Aleppo and Ar-Raqqa (the non-convexity constraint).

2. For each of the eleven governorates inside the envelope, shoot a ray from the origin (the point where x- and y-axes meet) through its point. Mark the intersection of the ray with the envelope as its projection point[6].

For visual clarity, we have drawn the ray and projection point only for Hama Governorate (green line). It is obvious that the rays of the eleven governorates intersect with three of the four segments of the envelope. By historical accident, none sends its ray through the horizontal line from Aleppo to the y-axis.

---

[6] Note for DEA purists: In this exposition, the term "projection" has a purely geometric interpretation. Its usage in DEA literature as "a formula for improvement", in the context of slacks and excesses (Cooper, Seiford et al. 2007), does not concern us.

3. On the ray for a given governorate, measure A = the distance from the origin to its point. Measure B = the distance from the origin to the projection point.
4. Calculate the humanitarian priority score of the governorate as the ratio A/B, its relative distance vis-à-vis the envelope.

DEA software will perform rules 1 through 4 automatically (some will do so implicitly, i.e. without reporting the projection points). The geometric interpretation is helpful for the understanding of the method as well as for a quick-and-dirty back-of-the-envelope calculation for a small set of units. From elementary triangle similarity it is obvious that it suffices to divide one of the measures by its projection point value. Thus, Hama has a severity score of approx. 0.06; its projection point measures approx. 0.22 on the x-axis. Its very quick-and-dirty priority score thus is 0.3. The program returned a score of 0.2817. Such precision is spurious; we have to assume considerable measurement error (and possibly also modelling error in the severity construct).

Governorate-level priority scores

This table collects the severity and magnitude measures and the resulting priority scores, sorted descendingly on the latter, as shown in the RAS report (SNAP 2015, op.cit., 37).

Table 1: Severity, magnitude and priority scores for 14 governorates

Governorate	Severity: Composite measure	Magnitude: Persons affected	Humanitarian priority score
Aleppo	0.18	2,824,000	1.00
Ar-Raqqa	0.26	741,000	1.00
Rural Damascus	0.22	2,135,500	1.00
Deir-ez-Zor	0.26	794,000	0.99
Dara	0.23	602,000	0.88
Quneitra	0.22	80,000	0.83
Idleb	0.19	1,428,000	0.83
Al-Hasakeh	0.09	632,000	0.37
Damascus	0.06	750,000	0.31
Hama	0.06	636,000	0.28
Homs	0.06	699,000	0.28
Lattakia	0.04	584,000	0.22
Tartous	0.03	278,000	0.14
As-Sweida	0.01	118,500	0.06

Several findings are of interest when we inspect both the graph and the table:

- Most noticeably (and notably), there is a distinct gap between seven high-priority governorates and seven of lesser priority.
- The relative influence of magnitude and severity on the priority score varies among governorates. Aleppo's is entirely determined by magnitude, Ar-Raqqa's entirely by severity. Rural Damascus is a mixed bag.
- In the benefit-of-the-doubt perspective, Quneitra and Dara are of particular interest. Their projection points lie on the vertical segment of the envelope. Therefore their priority scores depend only on their severity scores - magnitude has no influence. Dara has more than six times the number of affected persons, compared to Quneitra, but its priority score is only minimally bigger (0.88 instead of 0.83). It is bigger only because its severity score is also a tiny bit bigger.

This last finding raises philosophical issues - does Quneitra, with its small number of affected persons, deserve this high priority? Yes, as long as we accept the benefit-of-the-doubt reasoning, according to which the weight of indicators may vary[7].

Moreover, these priority scores hold only as long as the ensemble of the data is not changed in major degree. The scores are relative, as already noted. Suppose that Idleb's severity score were not 0.19, as calculated, but a high 0.30. If so, the envelope would be shifted. Aleppo would remain on it; Rural Damascus and Ar-Raqqa would lose their benchmark status, replaced solely by Idleb. Every other governorate's priority score would drop somewhat because all segments of the envelope (except the horizontal on top) would shift rightward, and with them the projection points.

The charm of the Syria case study lies in the visual explanation of the DEA logic. At this point, we should leave behind visual explanation. It works as long as we have only two variables ("outputs" in the DEA sense). Our next case study will attempt to extract priority scores from a model with *three* variables - measures not only of magnitude and intensity, but also of pre-existing conditions. At the stage, DEA involves an envelope that is not simply composed of lines, but of planes forming a three-dimensional structure. Adding further dimensions would produce hyper-planes, requiring algebraic explanations. The interested reader can easily find these on the Web, or in introductory books. Most readers will be more interested in substantive issues, and in the "do-it-yourself" of suitable DEA software.

[7] To assuage such philosophical questions, Checheye et al. (op. cit.) discuss setting minimum weights for each output. In our view, this would introduce a similar kind of arbitrariness for which classic indices with pre-fixed weights have been blamed.

Case study # 2: Philippines - comparing DEA to index results

Typhoon Yolanda struck the central Philippines in November 2013. The Protection Cluster collected detailed indicator data on 408 affected municipalities. Benini and Chataigner (2014) used these data to propose an index of unmet needs. The index multiplicatively aggregates three sub-indexes measuring pre-existing conditions, magnitude and intensity. The study and an Excel demo workbook, both available from the ACAPS Web site, document the underlying process and measurement models as well as the actual spreadsheet implementation. For convenience, the 2014 Excel demo workbook has been incorporated in one of the workbooks our demo files with this note.

The multiplicative model

Here we re-use the sub-indices within the DEA framework. We do not discuss how they were conceptualized and measured – the interested reader is referred to the 2014 study - but the difference in process models between the index of unmet needs then and what we are doing in DEA now is essential.

The multiplicative model was straightforward:

$$Needs = k * Magnitude * Intensity * f(Pre\text{-}existing\ conditions)$$

where *k* is an unknown constant expressing proportionality, and *f(.)* is a function of unknown shape and parameters, and * stands for the multiplication operator. Regarding *f(.),* we assumed that needs, given magnitude and intensity, were proportionate to the sub-index of pre-existing conditions (which combined poverty and malnutrition rates). Therefore, we used the untransformed sub-index of pre-existing conditions.

Two implications, however, need to be spelled out because they are not automatic in the DEA models:

1. Although the humanitarian and development needs of poor communities may be large, the index was crafted to signal needs only for communities impacted by the typhoon. The multiplicative formula guaranteed that unaffected municipalities had zero needs. This is so because zero magnitude or zero intensity would produce a zero value in the needs index.
2. In the index of unmet needs, all three sub-indices are so formulated that higher values express less desirable states. Higher magnitude is worse than lower magnitude, higher poverty and malnutrition are worse than lower levels, etc. Therefore the levels of unmet needs are plausibly estimated as the compounded effect of all three, by simple multiplication of the untransformed values.

Rethinking the model for DEA

The "benefit-of-the-doubt" logic in DEA obliges us to rethink this model. It seems desirable that, as in the previous index of unmet needs, very low-impact communities should receive relatively low index values even if their pre-existing conditions were already severe. However, if we simply gave the sub-index of pre-existing conditions (PEC) the same standing as the magnitude and intensity measures, the DEA algorithm would not meet this requirement. It would, literally, give the benefit of the doubt to very poor communities barely affected by the typhoon and assign them a high priority score by virtue of their extreme poverty[8].

Formally, this inequality holds:

DEA priority score for municipality X
$\geq \max\,[magnitude_X / \max(magnitude),\ intensity_X / \max(intensity),\ PEC_X / \max(PEC)]$

where max(.) is the maximum of the sub-index in the set of municipalities. The relationship is an inequality because the score can be higher than the maximum of the three ratios depending on how close the point is to the three-dimensional envelope.

By implication, the municipality with the highest PEC value receives a priority score of 1 – the highest in the DEA metric. Any municipality with a PEC value close to the maximum will score high, regardless of its magnitude and intensity values. This is not a desirable property for a needs assessment formula in a disaster relief context.

So, how can we change this? In a purely formalistic way, one could estimate the above model, followed by adjustments using restricted weights. DEA calculates variable weights on the same variables across units, depending on the segments of the envelope onto which a unit is projected. The admissible weights on a variable can be capped. Some authors recommend this or similar constraints for DEA models of social welfare indices. As noted on page 20, weight restrictions re-introduce the kind of arbitrariness for which the weighted additive index model is regularly blamed. We discourage them.

Substantive reformulations are more fruitful. They help us to so reformulate the process model that the DEA produces priority scores with desirable properties. In the Typhoon Yolanda scenario, we make the following assumptions:

- The PEC sub-index, formed of the poverty and malnutrition rates, is a vulnerability measure.

[8] Yet, on social justice grounds one might want to include them. See the footnote on page 40.

- Given the levels of magnitude and intensity, a municipality will recover the better the less vulnerable it was to this disaster. The less vulnerable, the more resilient.

The question then is how to redefine the PEC scores as *resilience* scores. For example, one may argue that resilience is inversely proportionate to vulnerability. Other transformations too would be possible. One could focus on positive resources for the recovery, as expressed through the proportions of the non-poor and the well-nourished. In other words, the proportion of people with financial and health reserves who are in positions to re-energize their communities. Obviously, different formulations of the resilience measure will produce different DEA results. The uncertainty about the best transformation reminds us that the DEA results always depend on model assumptions. We will return to this robustness issue. Here for the sake of demonstration we opt for

*Resilience score = k * (1 / PEC score),*

where k is an arbitrary scaling constant (DEA is indifferent to it).

Inputs and outputs

DEA originally was used to measure economic efficiency. It takes some conceptual jumps to understand how the efficiency algorithm can translate humanitarian priority. Efficiency internally is computed as the ratio of a weighted sum of outputs to a weighted sum of inputs, or, if you like, resources produced to resources consumed. The weights are flexible from unit to unit (e.g. from firm to firm). They are optimized for each unit so as to maximize its efficiency score, but the optimization is constrained by the rule that if the weights chosen for unit X were applied to the inputs and outputs of any of the other units, the ratio would not exceed 1.

Translated to the world of crises and disasters, resilience becomes an input, and magnitude and intensity become outputs. The efficiency score becomes the priority score. The unit (e.g. municipality) with the highest ratio of (weighted magnitude + weighted intensity) to resilience receives the highest priority (the denominator has only one term – resilience – so a weight for it is not needed). As in the economic model, the weights on magnitude and intensity can vary from municipality to municipality (and will do so for most). And, as emphasized, they are subject to constraints, which we demonstrate in this table.

Table 2: Hypothetical weights example (2 sub-tables)

Thus, in a contrived example of two municipalities A and B with

Municip.	Resilience		Magnitude		Intensity		Priority score
	Value	(not weighted)	Value	Individual weights	Value	Individual weights	
A	50	1	20	0.417	50	0.833	1.000
B	20	1	40	0.444	10	0.222	1.000

the weights were chosen such that both A and B attain the maximum priority score[9].

However, some of the weights for A are illegal. For, when we swap the weights between A and B as in

Municip.	Resilience		Magnitude		Intensity		Priority score
	Value	(not weighted)	Value	Individual weights	Value	Individual weights	
A	50	1	20	0.444	50	0.222	0.400
B	20	1	40	0.417	10	0.833	1.250

the priority score for B exceeds 1. Thus, the weights in A have to be lowered (e.g., by dividing the initial ones by 1.25). This implies that A must have a lower priority score than B. The DEA algorithm ensures that the weights are optimized such that they will not produce priority scores > 1 when swapped to any other municipality[10].

We will now report the results of this DEA specification of priority scores for the Typhoon Yolanda affected communities. Next we will compare their distribution to that of the multiplicative index in Benini Chataigner (2014, op.cit.). More to the point of this note, we will then walk the reader through running this model in a particular freeware DEA software.

Very different, yet same highest priority

Among the 408 municipalities, 370 had complete, positive values in the three variables. For these, the DEA application computed priority scores, with a range from 0.007 to 1, a median of 0.17, and a mean of 0.25. Only two communities attain the maximum score; i.e., only two lie on the envelope. They are different in an interesting way.

---

[9] In a spreadsheet, =(RC[-4]*RC[-3]+RC[-2]*RC[-1])/(RC[-6]*RC[-5]) will produce these results.
[10] For the technicalities, Cooper et.al. (2007, op.cit.) is very detailed.

Table 3: Comparing the two benchmark communities

ID	Municipality	Population 2013 (estimated)	Input	Outputs		Priority score
			Resilience score	Magnitude score	Intensity score	
321	PASTRANA	17,174	9.24	5.58	88.80	1.00
327	TACLOBAN CITY (CAPITAL)	228,147	24.02	100.00	100.00	1.00

Tacloban City was the hardest hit, relatively, in terms both of magnitude and intensity. In terms of resilience, the City is roughly at the 66th percentile of the 370 municipalities. So why would Pastrana, a small community with almost 20 times fewer affected people (magnitude score = 5.58), equally qualify for the highest priority?

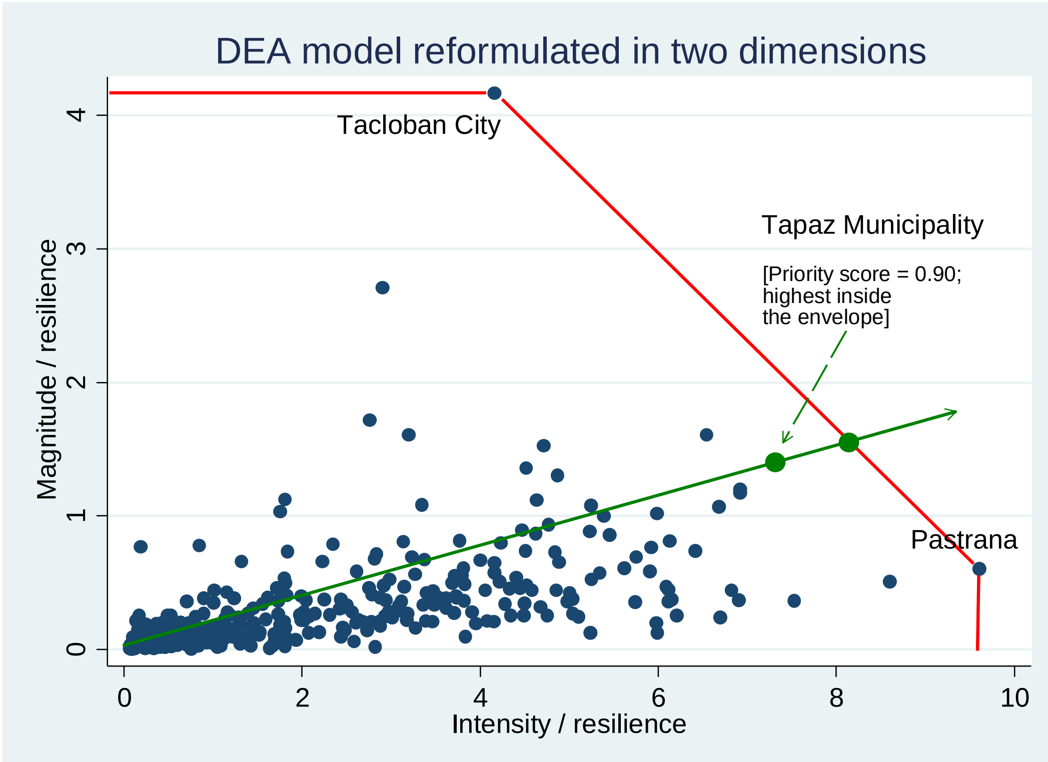
The cause is hidden in its low resilience, i.e. very high vulnerability. Its resilience score is at a very low 3rd percentile, i.e. 97 percent of the municipalities were less vulnerable. This induced the DEA algorithm to *"give Pastrana the benefit of the doubt"* and elevate it to the same priority as the much larger, much more affected Tacloban City.

Reduced and visualized

It is worth considering why that happened so that we need not just blindly believe the outcome of some black box mechanism. As mentioned, DEA computes efficiency scores (in our context the priority scores) as the ratio of the weighted sum of outputs to the weighted sum of inputs. In our Typhoon Yolanda model, there is only one input variable, the resilience score. Therefore this model is logically equivalent to a model with only two outputs and no input if we redefine the outputs as (magnitude / resilience) and (intensity / resilience)[11]. Such a model, however, can be visualized in two dimensions. This graph does just that. As in the Syria case study, the envelope is drawn red, and the sample projection green.

[11] And empirically as well. A test that we conducted in an abundance of caution confirmed this. "No input" means, technically, a dummy input = 1 everywhere. See section on OS-DEA.

Figure 6: Visualizing 370 typhoon-affected communities in the reformulated model



Also it is obvious that the majority of the municipalities project onto the segment linking Tacloban City and Pastrana. This group counts 288. A minority (77) project onto the vertical line dropping from Pastrana (77), and only three municipalities project onto the horizontal linking Tacloban City to the y-axis. In other words, for the majority of cases, both ratios – intensity to resilience, magnitude to resilience – together determine their relative priority. The weights within each of these groups are the same for all cases:

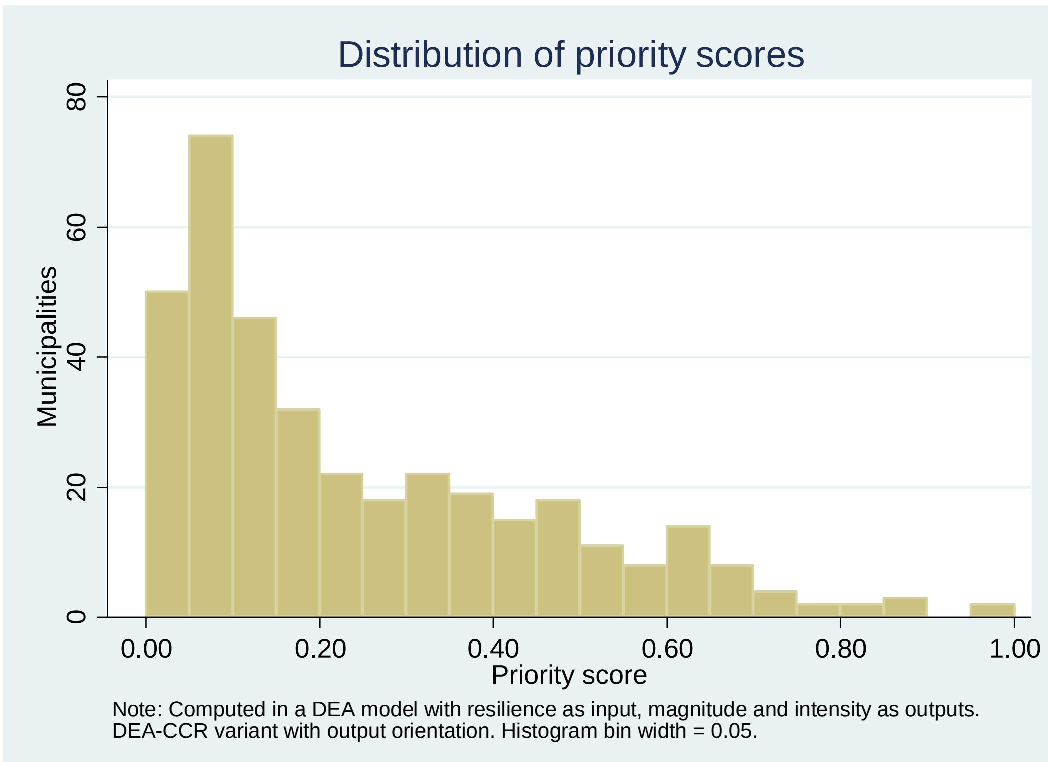
Table 4: Weights by segment of the envelope

Priority score depends on relative position to	Cases (inside envelope)	Weights on the ratios:	
		Intensity / resilience	Magnitude / resilience
Tacloban City only	3	0.000	0.240
Both TC and Pastrana	287	0.095	0.145
Pastrana only	78	0.104	0.000
Total	368		

The overall distribution

We return to the results of the one-input two-outputs model. The histogram suggests that the chosen DEA model produced a distribution characterized by few municipalities with high priority scores, more with middling scores, and a majority with low scores. From a policy perspective – well discriminating priorities -, this seems desirable.

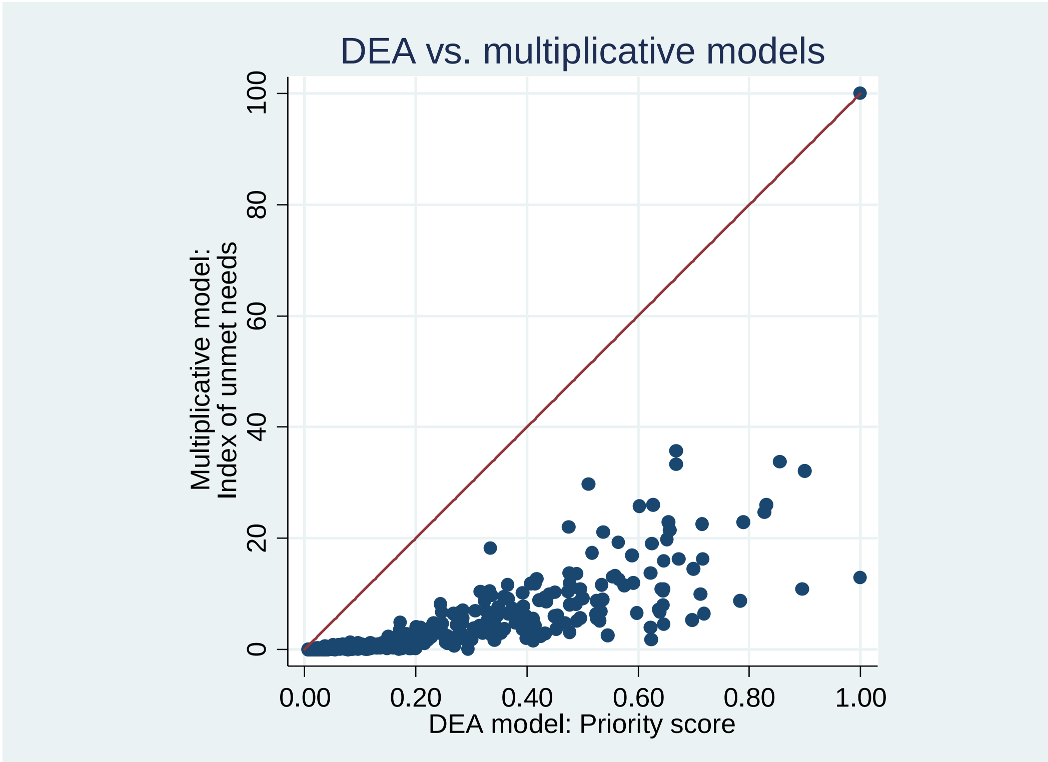
Figure 7: Distribution of the priority scores in the Typhoon Yolanda model



Correlation between the multiplicative and the DEA models

Our key interest is to compare the pattern of severity scores produce under the DEA with that known from the multiplicative models. How are they correlated? In particular, how are the ones assigned the highest priority by one method faring under the other method?

Figure 8: Comparison of the DEA and multiplicative models



The differences are major. The distribution in the multiplicative model is notable for the large difference between the no. 1 – priority municipality (Tacloban City with the maximum index value scaled to 100) and all the others. The DEA approach makes for a much closer leading pack – not only are there two municipalities in the first rank, there are five others with scores higher than 0.8. It is the "benefit of the doubt" design that produces this more even field.

The two indices are highly correlated (Spearman's rank order correlation = 0.92), but this is driven mostly by the mass of cases clustering in the low ranges of both indices. At higher ranges, the correlation weakens. For example, for the subset with DEA priority scores > 0.60, Spearman's drops to a mere 0.36. In other words, the ordering of high-priority municipalities differs sharply.

Of greater interest is the question which of the three factors – vulnerability (or, in the DEA model, resilience), magnitude and intensity – drive these differences most strongly. For the multiplicatively constructed index of unmet needs, this is relatively straightforward. The model is fully additive in the logarithms, and the contributions thus depend only on the coefficients of variation in the factors. The analysis of variance – not shown here for space reasons – reveals that vulnerability accounts for 16 percent of the variation, magnitude for 31 percent, and intensity for 53 percent.

That decomposition is based on the entire set of 370 affected municipalities. Policy-wise that is not very relevant. We want to know what differentiates those that the multiplicative models pushes to the higher ranks. For comparability we look at the 35 municipalities only that scored higher than 0.60 in the DEA model. Then the influences change dramatically: the multiplicative model is almost entire driven by differences in magnitude. The rank-order correlation between the index and the population is a blistering 0.96. Vulnerability and intensity matter little.

The influence structure in the DEA is more balanced. For better comparability, we study the impact of vulnerability, not resilience. For the same subset of 35 municipalities, vulnerability, magnitude and intensity account for 63 percent of the priority score variance[12]. Their respective shares are 21, 15 and 28 percent. The rank-order correlation between the priority score and the population is a mere 0.21.

Is a balanced index necessarily better?

For needs assessments that wish to give similar consideration to several priority factors – such as vulnerability (pre-existing conditions), magnitude (number of affected persons) and intensity (proportion of impacted units) – a formula that balances their influence seems to be a matter of fairness. Therefore DEA, as far as the Philippines results tell, seems intrinsically preferable to methods that privilege one factor at the expense of the others.

However, on policy grounds that is not necessarily always correct. The way to evaluate the fairness of prioritization formulas is to take a close and detailed look at the distributions of every factor considered and measured. If any of them takes high or even extreme values for a significant group of cases, and most members of this group have more or less average values on the other factors, we should accept that their priority scores are driven chiefly by this one distinguishing factor. Such a scenario is not implausible in regions dominated by some larger cities amid a host of smaller rural communities, in a disaster or crisis that affects many of them with similar intensity.

In this scenario, the multiplicative model will identify all or most of these larger communities as top priorities. The differences in intensity are too small to lift the priority scores for smaller communities, even the relatively more strongly impacted ones, to the upper

[12] With all variables log-transformed, as in the decomposition of the multiplicative unmet-needs index. In the DEA model, the priority score is not an additive function, regardless of what transformations are used. The logarithmic transformations are chosen here only for comparability with the multiplicative model, which does have additivity in the logarithms.

ranges. In terms of allocating relief, the focus on larger communities may then seem justified.

What DEA does in this situation is to emphasize that our priorities may be based on one-sided considerations. Communities for which we calculated a relatively high disaster intensity (as a proportion of the maximum intensity score) should also be prioritized, regardless of their size. Relative values matter, that is to say the effects on priorities should result from both the model and the data, rather than being hard-wired in advance.

Whether one prefers one or the other position will ultimately be decided on philosophical and policy grounds. Here, all we need to note is that DEA is data-driven, but not entirely so. It is equally strongly based on model choices. In this example from the Philippines, we made a distinct choice by the way we treated vulnerability. Instead of placing in on an equal footing with magnitude and intensity, we transformed it to a resilience score to be entered on the DEA input side. We did so in order to prevent highly vulnerable, but barely affected municipalities to attain high priority scores. The cost of this manipulation – apart from the difficulty of explaining it – is that it forces inflexible relationships between vulnerability on the one hand and magnitude and intensity on the other. We note this point to avoid the impression that model-driven approaches (e.g., the multiplicative index) can be replaced by an entirely data-driven method. DEA is model *and* data-driven.

Computing the priority scores with OS-DEA

General notes on this freeware application

There is a considerable variety of available DEA applications. The Web carries review sites for them; older reviews, however, quickly grow obsolete in this mutating field. Of particular interest are applications that team with the workhorse of humanitarian data analysis, MS Excel.

We calculated the priority scores of the Syria and Philippines DEA models using the freeware application "OpenSource DEA"[13]. OS-DEA is a suitable choice for our purpose. It has several attractive traits as well as some downsides. We note both briefly, before demonstrating the process of computing the priority scores of the Philippines model.

Pros

- **Importing the data:** OS-DEA imports comma-delimited (.csv) files, a format in which Excel can save data tables. Variables are clicked into input and output boxes. Results are inspected in different views. DEA expects at least one input variable. In order to run a "pure output" model (the Syria situation), a dummy

[13] http://www.opensourcedea.org/index.php?title=Open_Source_DEA.

variable with uniform value 1 (to be created in Excel) conveniently fills this requirement. In mixed input-output models (e.g., the Philippines case), the input and output variables(s) are easily segregated in the set-up process within OS-DEA.

- **Exporting to Excel:** OS-DEA's most appealing feature is its ability to export data, internal parameters and results to a neatly structured Excel workbook. Separate sheets record the model details, raw data, efficiency scores ("priority scores" in our context), projection point coordinates, the names of the nearest units on the envelope (the so-called "peer group"), the unit-specific weights (the priority score is the ratio of weighted outputs to weighted inputs), as well as some parameters of minor interest.
- **Fast calculation:** Driven by a Java linear programming machine, the computation of scores, even for hundreds of units, takes a few seconds only.

Cons

- **Models and calculation:** The application offers a modest number of different DEA models to choose from. In the original context of economic efficiency measurement, this versatility is a strength. In the social indicator field, these choices may confuse the user; most models will lead to inappropriate or hard-to-interpret results. We strongly advise using *only* the "CCR_O" choice[14] (Model choices are set in a drop-down menu that offers a nutshell description of each model).

- **Limitations:**
    - Typically of freeware, the documentation is sparse. It is barely sufficient to get the user started.
    - We found the most difficult aspect to be the installation. The choice of .dll files from the downloaded zip file and their destinations depend on the operating system as well as on the Java 32-bit or 64-bit flavor. Some trial-and-error experimentation may be required before OS-DEA works.
    - Records with missing or zero values are rejected; the model does not execute, and the cleaning has to be done back in Excel, and the cleaned

[14] The acronym stands for the so-called Charnes, Cooper and Rhodes Model with output orientation. This model was first introduced in 1978. It assumes constant returns to scale. While "returns to scale" is not equivalently translatable to the social indicator context, its complement "variable returns to scale" would produce results that are difficult to interpret. Whether the model should be input-oriented (as Cherchye et.al. 2007, op.cit., recommend) or output-oriented, is of no concern (as Cooper et al., op.cit.: 115, themselves attest). Since the "pure output" models reflect a frequent situation in our context, with a fixed dummy input variable, the idea of reducing input amounts seems less appropriate. We prefer, and have used, output-oriented models.

file re-imported. Apart from being inefficient, this rules out units that have legitimate zeros in one or several of their output variables.

- Values of the imported data smaller than 0.001 are correctly handled, but are displayed as zeros, an irritation.
- OS-DEA does not offer robustness tests. However, limited checks can be done manually, by running the model again on one or several datasets from which a purposive subset (usually the envelope members of the previous model) or a small random sample of units have been excluded[15].
- OS-DEA handles models with several input and several output variables. Therefore it returns the computed flexible weights in each variable and for each case (the essence of DEA!). In pure-output or in one-input models, the input weights, as returned, are a distraction. The output weights should be normalized (by dividing them by the input weights, and all input weights become 1). This does not concern users who want to know only the priority scores. For those wanting to work with the weights applied to specific cases (and to see that the normalized output weights are actually identical among all cases projecting to the same envelope segment!), this has to be done manually after exporting results to Excel.
- The OS-DEA does not appear to be actively maintained at present. There is an option of exporting results in the .xlsx format, which arrived with Excel 2007. There is no user support. The download supports 32- as well as 64-bit machines and Java flavors.

While we have used OS-DEA effectively, we do not want to prevent readers from exploring and working with other DEA applications. We do recommend that persons seriously interested in DEA consult an introductory text beyond the OS-DEA documentation.

Walking through the process

OS-DEA installation

The Web-based installation instructions are helpful except for one misleading sentence. They make it appear that only one of the two .dll files is needed. In fact, *both* must be copied to the designated directory. Java needs to be pre-installed (most machines will already have it). The Java version (32 or 64 bit) needs to be known; upon it depends the correct choice of the .dll file setting to be copied. If this remains confusing, some experimentation with different sets copied to different locations may be helpful. OS-DEA will tell only in the third step of building the model whether it finds the required files or not. Since these are copied to the Windows system, the user needs administrator privileges or

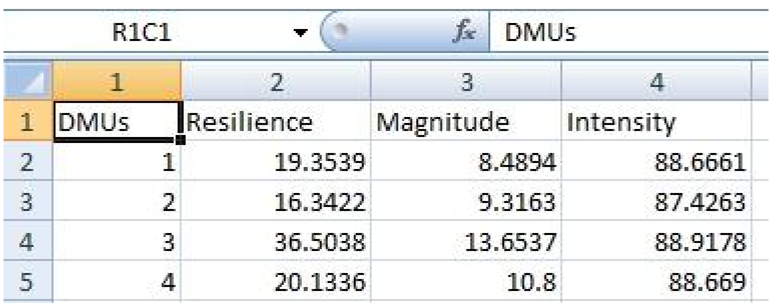
[15] It appears that only high-end commercial DEA applications offer built-in robustness checks.

support. Other than for the .dll files with their prescribed location, the locations of the executable “OSDEA-GUI-v0.2.jar” and of any data files involved can be freely chosen.

Creating and saving the data file

In Excel, we create a file containing only the variables that are going to be used in OS-DEA. If additional, non-used variables are included, OS-DEA will import the data, but will refuse to finalize the model. The leftmost column must contain the unit identifier. It can be named freely. In keeping with DEA tradition, we named ours “DMUs” in order to mention this acronym at least once. DMUs are “Decision making units”, commonly economic agents who decide the allocation of inputs (and monitor the outputs). In most non-economic contexts “Record ID” or simply “ID” would be appropriate. Importantly, the ID variable must, as usual, have unique values; it should be the ID variable later used to link back (via VLOOKUP if necessary) the results to the Excel datasheet in which they will be further analyzed.

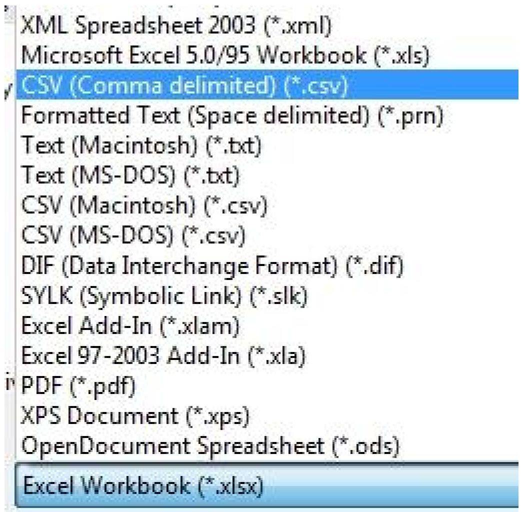
Figure 9: Screenshots of using the OS-DEA application


R1C1 | fx DMUs

	1	2	3	4
1	DMUs	Resilience	Magnitude	Intensity
2	1	19.3539	8.4894	88.6661
3	2	16.3422	9.3163	87.4263
4	3	36.5038	13.6537	88.9178
5	4	20.1336	10.8	88.669

Our experience is that OS-DEA does not accept datasets with records that have missing values or zeros. Thus such records must be purged; all variables other than DMUs must be numeric and positive. An alternative is to assign small positive values in lieu of zeros if this can be justified (certainly for legitimate zeros on the output side. This is less likely so on the input side, where small values are liable to lead to abstruse values in the output sum / input sum ratio) or to impute missing with median or model-derived values.

The data must be saved in .csv (comma-delimited) format. It is recommended to inspect it in a text editor; for unknown reasons, Excel occasionally terminates lines with commas. The lines have to end in the values of the rightmost variable, without punctuation (which is the correct .csv formatting anyway).

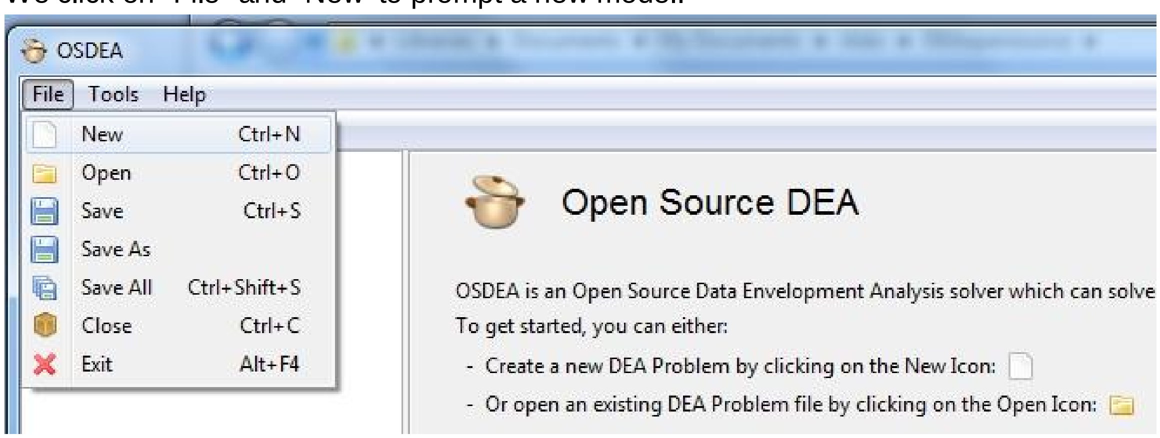


Treatment within OS-DEA

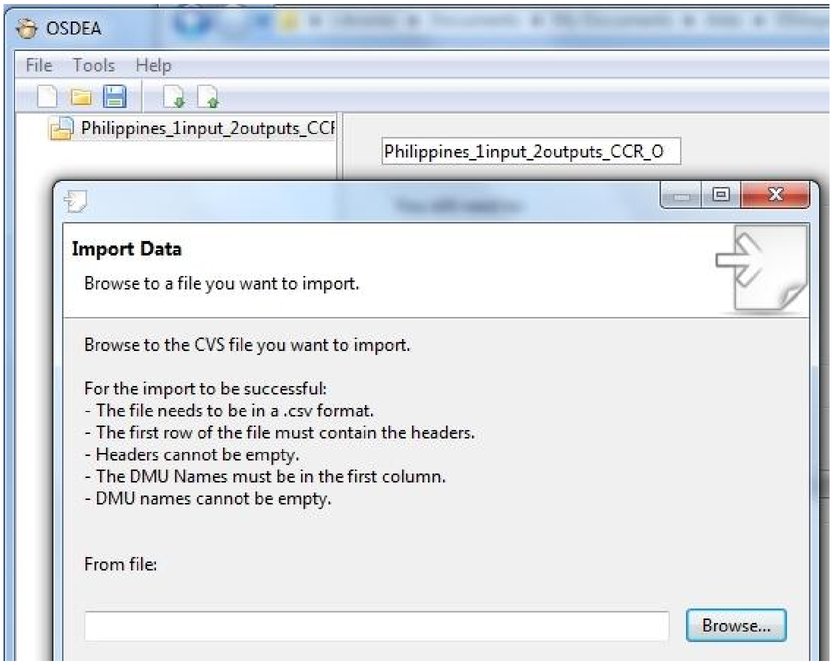
For this demonstration we have used, in the demo zip-file, the .csv file *"150707_Input_File_for_OSDEA.csv"*. The reader is welcome to use his/her own; the steps should be clear from the following instructions and screenshots.

Starting a new model

We click on "File" and "New' to prompt a new model.



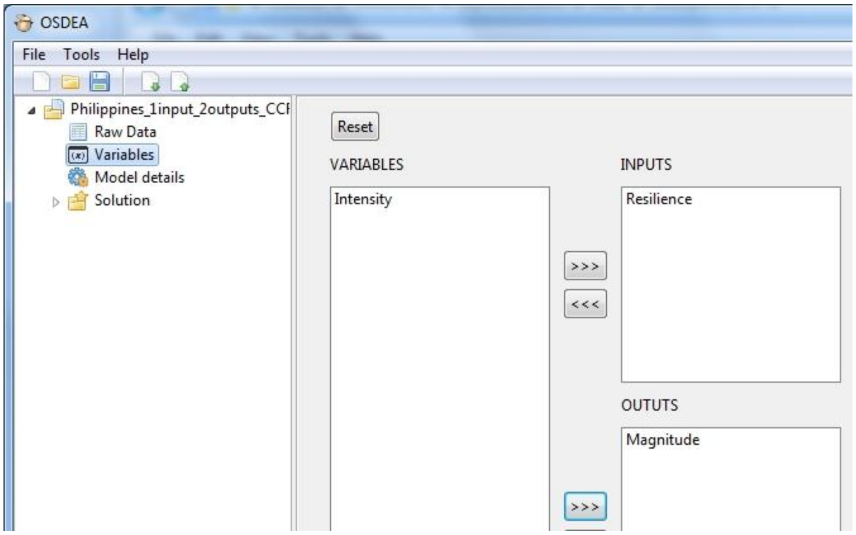
Naming the model and importing data



Optionally, we can give the model a name, by typing in the box for model name above to the right. Then we click Tools and Import and browse to the .csv file to be opened.

Assigning the variables to inputs and outputs

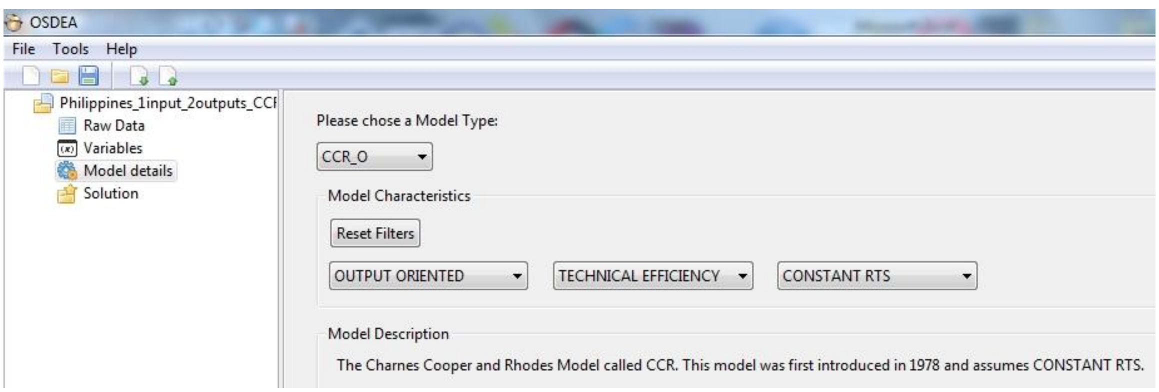
We click on the named model icon to expand it. Optionally, we can inspect the Raw Data, but cannot edit them in OS-DEA. We click on the Variables icon; a surface opens to let us click the variables to their assigned use as inputs and outputs. In this screenshot, “Resilience” and “Magnitude” have already been assigned. “Intensity” awaits being clicked to the Outputs box.



We do not use the boxes for non-discretionary and non-controllable inputs or outputs (not shown in this screenshot).

Choosing a model type

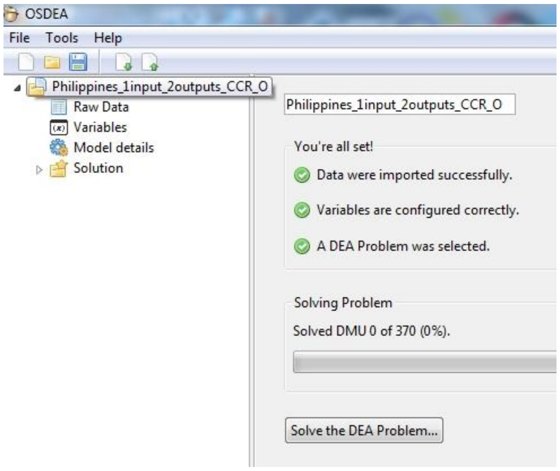
We click the "Model details" icon. On the right side, we make sure to choose "CCR_O" as the model type.



We do not change the choices displayed in the other drop-down boxes.

Calculating the model solution

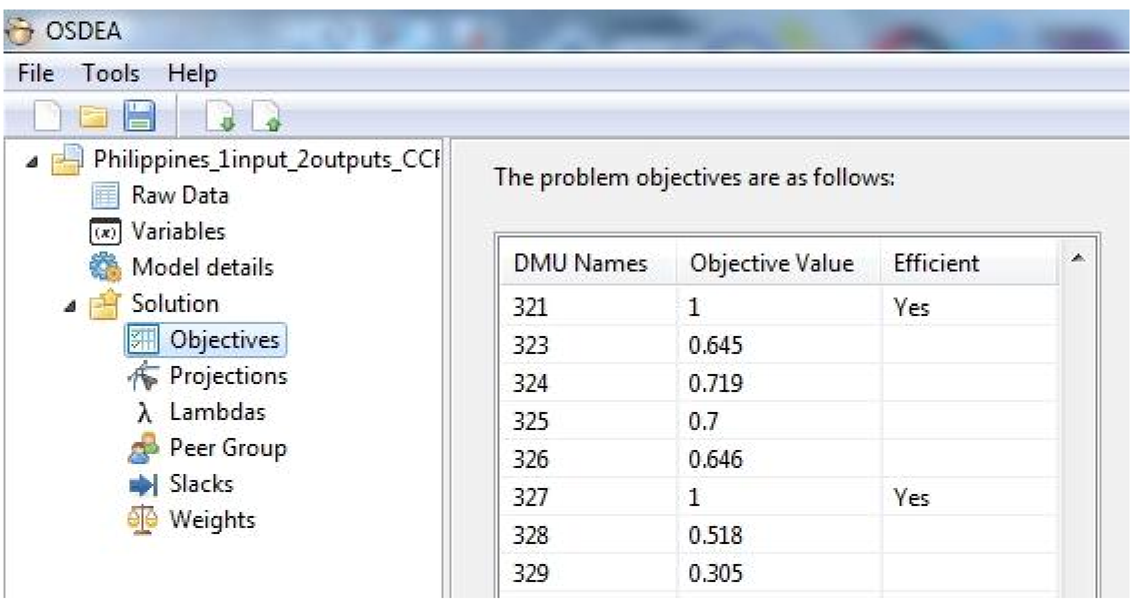
This is the most confusing part, liable to frustrate neophytes. Clicking the "Solution" icon goes nowhere. We must go back to the root and click the model name icon again.



This will bring up, on the right side, a report that all is set and ready to solve the DEA model. We click the "Solve the DEA Problem…" button.

Inspecting the solution

After successful execution, the "Solution" can be expanded. Five items appear. In our context, "Objectives", "Peer Group" and, for some users, "Weights" matter.



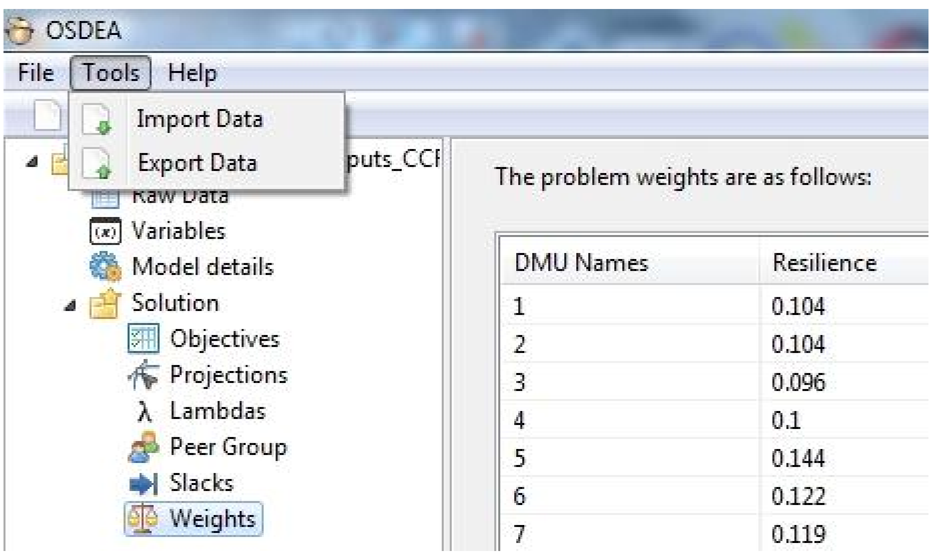
Our key interest is in the priority scores. We find them in the column "Objective Value" (the lingo is from efficiency analysis). Members of the envelope are shown with a "Yes" in the "Efficient" column. This model has exactly two units with priority score = 1. As we already know, these are the municipalities of Tacloban City and Pastrana mentioned in the previous section.

The “Peer Group” variable lets us see on which segment of the envelope the unit was projected, in other words whether its priority score depends on its position vis-à-vis just one or several of the envelope members. This is a convenient result to have. Later, in Excel, a Pivot table can quickly produce counts of units (here: municipalities) whose priority scores are determined by their magnitude to resilience or their intensity to resilience ratios or by both. A severe imbalance – low counts in any of the three categories – should lead to case-by-case inspection in the search for outliers and, depending on findings, special treatment of these (exclusion and re-estimation of the model without them).

Some users may want to analyze the weights. The remarks on normalizing weights in the case of pure-output or one-input models pertain. Lambdas and slacks are constructs used in efficiency measurement contexts; they are not useful in ours. Projections are interpretable only in the simplest model – no input, two outputs. Here they define the coordinates of the projections on the envelope segment lines.

Exporting the results to Excel

Finally, we click “Tools” and “Export Data” and save the results in an .xls or .xlsx file.



The Excel workbook will have these sheets:

- Model details
- Raw Data
- Variables
- Objectives

- Projections
- Lambdas
- Peer Group
- Slacks
- Weights

Further processing in Excel

These steps seem the most productive when we integrate the DEA results with the workbook from which the data set for OS-DEA was taken:

1. In the exported workbook, combine the variables from the sheets Objectives, Peer Group and Weights in one sheet.
2. Rename the variables in the spirit of priority scores, e.g. “Objective Value” as “Priority score”, “Efficient” as “Envelope member”. The weights variables have to be pre-fixed with “weight” or “w_”, e.g. “w_Resilience”, in order to avoid confusion with the original data variables.
3. Name this table conveniently, e.g. as “CombinedResults”, and copy this table to the original workbook. If no cases were deleted for the OS-DEA .csv datafile, and the sort order is the same, the new variables can simply be copied to the original data table. If cases were deleted, use the look-up function VLOOKUP to merge the data.

Further analysis then proceeds as needed.

General issues in using DEA

Beyond the pros and cons, facilities and limitations discussed in the previous section, users of DEA in priority scoring models need to be aware of some recurrent challenges. Here we discuss three of them. The first is generic, apart from the choice of DEA as a method to calculate priority scores. #2 and 3 are inherent of the DEA logic. None is serious enough to discourage exploring the DEA potential for this purpose.

Issue #1: Treatment of the pre-existing conditions variables

Ideally, all the sub-indices – pre-existing conditions (PEC), magnitude and intensity – would be entered as DEA outputs, and the model would be run as a pure-output. This would, analytically, place all three on an equal footing; there would, at first glance at least, be no need to justify different treatment, as inputs vs. as outputs.

However, as we explained earlier, this would imply an incorrect process model. By the “benefit of the doubt” logic in DEA, the community with the highest pre-existing condition score would automatically attain the highest priority score, regardless of how low its magnitude and intensity scores were. Similarly, other communities with high PEC scores would receive priority scores at least as high as the ratio of their PEC scores to the highest such score in the set, apart from all other considerations.

If this is not desirable, then the PEC sub-index has to be treated differently[16]. For the Typhoon Yolanda model, we at first interpreted it as a vulnerability measure. This far many analysts might follow us, accepting such an interpretation for contexts with available pre-disaster indicators. Conceptual demands become more ambitious, and potentially more contentious, when we, in a second interpretation, assume that resilience is inversely proportionate to the PEC level. The reinterpretation is needed in order to enter PEC, in new guise, on the input side. DEA puts the weighted sum of inputs into the denominator; since more vulnerable communities should receive higher priority (other things being equal), its reciprocal – the resilience measure – has to go there.

If one follows us to this point, another objection arises. Logically, our Philippines model is equivalent to a model with no input and these two outputs: 1. the product of magnitude and PEC scores; 2. the product of intensity and PEC scores. This follows from the simple identity 1 / (1/x) = x.

This objection is correct to the point that the two multiplicative terms firmly link the individual sub-index values by an arithmetic operation; they are not dependent on the ensemble to the data. The model thus becomes a hybrid between DEA and the multiplicative index. There is no “benefit of the doubt” within the multiplicative terms, only between them.

But this argument does not prevail. There are two benefits to the kind of input/output model tried out on the Typhoon Yolanda data:

- First, resilience may not be the only variable that the analyst wishes to place on the input side.

There may also be other variables that should determine priority. *Humanitarian access* may be one of them. Suppose that after an earthquake in a mountainous region (think of Nepal in 2015!) some communities find themselves cut off from the main trunk road due to landslides and collapsed bridges. The population living upstream from the lowest obstacle is a measure, or part of a measure, of humanitarian access, at least until alternatives have been opened. The analyst may not want to simply factor this information

[16] Whether this indeed is not desirable is a question of social justice concepts, not of DEA modeling. Some might make a case for including all three sub-indices on an equal footing, on the grounds that chronic (structural) and acute (crisis-related) deprivation should be addressed equally. Most will likely point to political realities, in the sense that national and international actors will focus the response on the populations clearly affected by the crisis, rather than assisting all highly deprived groups regardless of causes.

into the magnitude or intensity measure. Similarly, as a short-term state of affairs, the distribution of these obstacles should not be mixed with the measure of long-term resilience. Thus, in the DEA logic, the proportion of people in the district NOT cut off could be used as a second input variable. The denominator would then consist of the sum of the (flexibly weighted) resilience and "still accessible population" measures.

In other words, the input / output model is more flexible to accommodate situations with data that are not simply reducible to PEC, magnitude and intensity.

- The second, and stronger, defense is to say that a valid process model cannot be captive to an analytic algorithm.

If our understanding of priorities demands that certain concepts work together differently from the pure-output setup in DEA, then these links have to be made. Not all simple models are valid models. Ultimately, validity trumps simplicity.

Issue #2: Treatment of cases with zeros

Analytically, several DEA models – including the CCR_O model that we used and recommend – permit semi-positive inputs and free outputs (Cooper et al., op.cit., 114-5). This means that every unit needs at least one input > 0, and all other inputs ≥ 0. This conditions prevents division by zero. The output measures are free, i.e. they can take negative, zero, or positive values.

OS-DEA, in our experience, aborts the calculation whenever it encounters a zero values. We have not yet fully tested this, but we found that for estimating the Typhoon Yolanda model, we had to delete records with zeros. This was barely more than an inconvenience. We only had to exclude 32 non-affected communities. In other contexts and data situations, this restriction may be a source of bias. It may be acceptable to replace zeros with very small positive values, but this has not been tested.

Issue #3: The so-called "modifiable areal unit problem"

This is the name given to the paradox that relationships between variables can be substantially altered depending on the way the data are aggregated. It is well known in quantitative geography (Fotheringham, Brunsdon et al. 2000)[17].

To illustrate this, we return to the Syria model. We consider the relationship between a spatial variable - the governorates - and the priority scores. The scores differ considerably depending whether they were computed directly off the governorate-level variables, or first

[17] Further references at http://en.wikipedia.org/wiki/Modifiable_areal_unit_problem.

for the 270 sub-districts and then aggregated. This table compares them; population-weighting in the rightmost column was by pre-crisis population.

Table 5: Governorate-level priority scores, by aggregation mode

Governorate	14 governorates: Governorate scores	270 sub-districts: Sub-district scores, mean by governorate	
		Unweighted	Population-weighted
Aleppo	1.00	0.40	0.74
Rural Damascus	1.00	0.38	0.40
Ar-Raqqa	1.00	0.37	0.46
Deir-ez-Zor	0.99	0.33	0.41
Dara	0.88	0.36	0.36
Quneitra	0.83	0.33	0.40
Idleb	0.83	0.27	0.30
Al-Hasakeh	0.37	0.24	0.22
Damascus	0.31	0.92	0.92
Hama	0.28	0.16	0.21
Homs	0.28	0.18	0.34
Lattakia	0.22	0.11	0.31
Tartous	0.14	0.06	0.08
As-Sweida	0.06	0.03	0.04

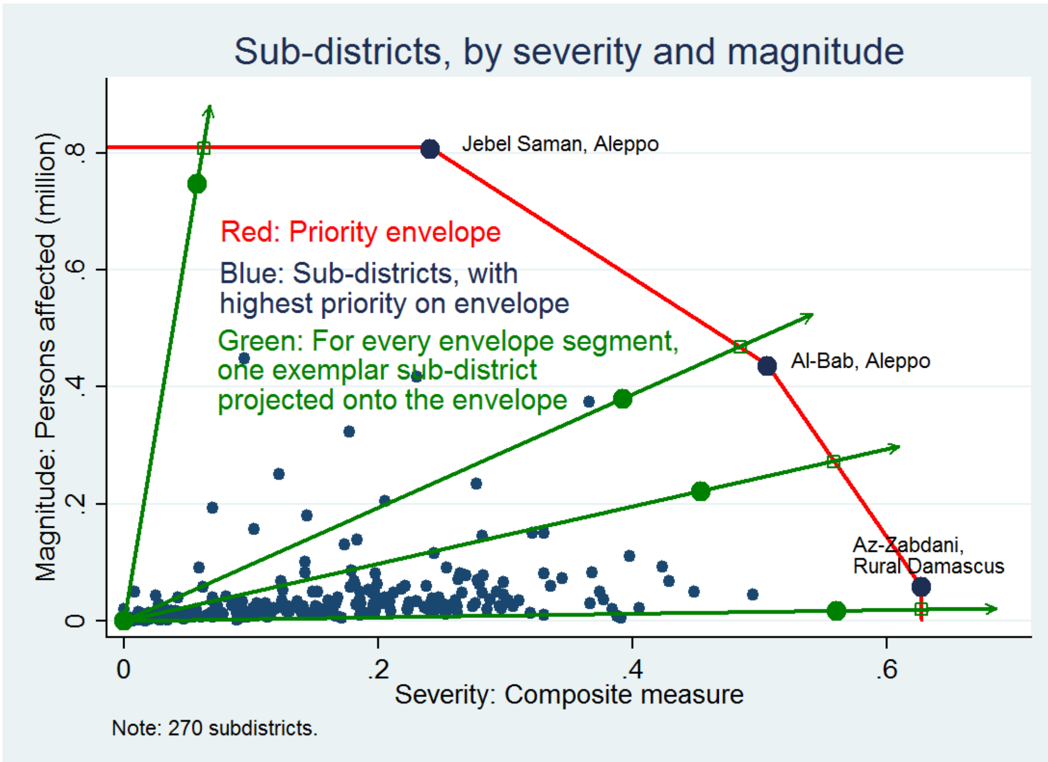
The differences are caused by two mechanisms. One is accidental to the administrative divisions of Syria and to the history of the conflict, the other inherent in the DEA logic.

1. **Syria:** Governorates have different numbers of sub-districts. Damascus has only one; Aleppo has 40. In terms of magnitude (persons affected), the relative differences among sub-districts are more pronounced than those among governorates[18]. Therefore, it is likely that from among the governorates with high numbers of affected persons, some sub-districts will have particularly high numbers relative to all others in Syria. If they are in a governorate with few sub-districts (Damascus!), they will push up the governorate mean score. The next figure repeats the basic geometric representation of the DEA model, this time for the sub-districts. The dot on the steepest of the four rays represents Damascus. It is projected onto the horizontal segment. Therefore its priority score depends *only* on the magnitude, not at all on the severity. The score is simply = affected persons in Damascus / affected person in the Jebel Saman district of Aleppo = 750,000 / 812,000 = 0.92. The governorate of Damascus

[18] The coefficient of variation in the numbers of affected persons is 0.87 for the 14 governorates and 2.0 for the 270 sub-districts.

inherits it fully from this one sub-district, also called Damascus - a purely Syrian peculiarity.

Figure 10: Scatterplot for the sub-district-level DEA



2. **DEA-inherent:** This problem arises also on the severity side of our Syria model, and in general on the intensity side. It is generic and will not go away if the large urban sub-districts are further sub-divided in DEA modeling. Take again the ratio of affected persons to pre-war population. Assume that the ratio is a randomly distributed variable in a set of administrative units. DEA results, such as our priority scores, are defined in terms of the relative distance to the envelope. If this ratio were used as the intensity measure, at least one segment of the envelope would be defined by its maximum. The expected maximum from a draw of 270 units is higher, possibly much higher, than from a draw of 14, regardless whether we are in Syria or elsewhere. In fact, in the Joint Data Review dataset, the maximum for the sub-districts is 5.4; for the governorates it is 0.78, as already noted. Inversely, the mean/max ratio is lower for the sub-districts than for the governorates. As a consequence, when aggregating to the governorates, the priority scores tend to be lower than when they were directly calculated from the governorate data. This is evident in Table 5 above.

This issue does not arise in classic index formation with fully additive indicators. In this case the population-weighted aggregated values equal those directly estimated at the higher level. DEA-generated indices lack this property. One may argue that we are not interested in resolving this inconsistency; our chief aim is to express humanitarian priorities at the highest geographic-administrative resolution possible, currently the sub-districts. However, dependency on some extreme values is likely to lead to findings that will be seen as unfair in policy terms even if they are formally correct. Finding appropriate treatments for extremes during indicator preparation should therefore be high on the agenda of DEA modelers.

Conclusion

DEA is an instructive example of conceptual migration. Invented to tackle a challenging apples and oranges problem in public policy, it found a place in the economists' toolbox. Economists have used it to measure the efficiency of firms and programs with key variables for which market prices did not exist. Hence DEA percolated to other domains, including, for the last ten years or so, to social indicator research. From this side (and also from the efficiency side via humanitarian logistics) it is knocking at the doors of humanitarian data analysis.

Here it is a candidate method in prioritization. Earlier in this note, we delineated its niche amid classic additive and more recently favored multiplicative index forms. DEA is at its best when analysts wish to hold off on *apriori* weights or aggregation functions. It is particularly intuitive when only two measures – one for the magnitude of the disaster or crisis, the other for its intensity – are to be considered. In this case, a two-dimensional visualization is possible and indeed instructive. When there are few cases – e.g. only the 14 governorates of Syria -, one can calculate acceptable approximations of the priority scores manually, on a print-out of the magnitude vs. intensity scatterplot. With a ruler one draws the envelope, and the projections, and measures distances from origin to points, respectively their projection points. One makes the divisions in one's head, or at most with the help of a hand calculator – and boom! we have a set of priority scores good enough for the moment. No software needed.

In more complex situations, we rely on DEA software. The advent of applications that are reasonably cheap (or even free), convenient (working in or with Excel) and efficient is the decisive factor why we are discussing this method at all. However, apart from the effort that it takes to master the software, the conceptual demands, too expand. We have seen this when we moved from the simple Syria case to the more involved Philippines case. Not only was the number of units much larger (402 municipalities vs. 14 governorates), the way the additional measure – the index of pre-existing conditions – had to be integrated was less than straightforward. At this level of complexity, DEA loses its immediate intuitive appeal. It demands analytic decisions that can no longer be defended as “simply data-driven”.

One possible consequence is to use DEA models as a temporary measure, in the early stages of priority index building. In this perspective, DEA is a discovery mechanism. It suggests starting values for priority scores on which subsequently analysts can improve, to the extent that they understand the mechanics of crisis impacts better and better. If they do, they might want to replace the initial DEA results with those obtained from updated additive or multiplicative index models. These are easier to compute (entirely within Excel) and easier to test for robustness (by tweaking weights or aggregation functions).

It is far from certain that in any given crisis we will, within useful time, arrive at valid impact measures based on classic indices, to the point where we can dispense with the “benefit of the doubt” approach inherent in DEA. The comparison of methods in the Philippines case study suggests that DEA produces a more evenly distributed priority score pattern. The multiplicative approach promotes one community – a provincial capital – to an extreme outlier position. There were reasons for this – notably the size of its population. But we also show that for the small group of high-priority municipalities, the multiplicative index is informed primarily by magnitude. By contrast, the DEA priority scores reflects magnitude, intensity and pre-existing conditions in a more balanced way.

We are at the very beginning of an experimental process. No far-reaching conclusions can be taken. All that we can recommend at this stage is a modest learning investment, in order to get going, and a comparative attitude in which intelligent questions arise when we compare results obtained with DEA to those of traditional methods. It will take several small experiments, followed by at least one major needs assessment using DEA prominently and successfully, before this method can claim parity with the established methods, however questionable these may be.

References

Alsharif, K., E. H. Feroz, A. Klemer and R. Raab (2008). "Governance of water supply systems in the Palestinian Territories: A data envelopment analysis approach to the management of water resources." Journal of Environmental Management 87(1): 80-94.

Benini, A. (2012). Composite measures. Their use in rapid needs assessments. Conceptual background and technical guidance. Geneva, Assessment Capacity Project (ACAPS).

Benini, A. and P. Chataigner (2014). Composite measures of local disaster impact - Lessons from Typhoon Yolanda, Philippines. Geneva, Assessment Capacity Project (ACAPS).

Cherchye, L., W. Moesen, N. Rogge and T. V. Puyenbroeck (2007). "An introduction to 'benefit of the doubt'composite indicators." Social Indicators Research 82(1): 111-145.

Cooper, W. W., L. M. Seiford and K. Tone (2007). Data envelopment analysis : a comprehensive text with models, applications, references and DEA-solver software. New York, Springer.

Decancq, K. and M. A. Lugo (2012). "Weights in Multidimensional Indices of Wellbeing: An Overview." Econometric Reviews 32(1): 7-34.

Fotheringham, A. S., C. Brunsdon and M. Charlton (2000). Quantitative geography : perspectives on spatial data analysis. London ; Thousand Oaks, Calif., Sage Publications.

Nardo, M., M. Saisana, A. Saltelli, S. Tarantola, A. Hoffman and E. Giovannini (2008). Handbook on Constructing Composite Indicators: Methodology and user guide. Paris and Ispra, OECD and the Joint Research Centre (JRC) of the European Commission.

Ravallion, M. (2010). Mashup Indices of Development [Policy Research working paper # 5432]. Washington DC, The World Bank.

SNAP. (2015). "Regional Analysis of the Syria conflict (RAS), Q4 - 1 October-31 December 2014, Part A.1., Overview and Sectoral Analysis." Retrieved 18 February 2015, from http://www.acaps.org/img/reports/p-regional-analysis-for-syria---part-a-overview-and-sectoral-analysis-oct-dec-2014.pdf.

UNOCHA. (2014). "2015 Humanitarian Needs Overview. Guidance: Humanitarian Needs Comparison Tool [August 2014]." Retrieved 17 February 2015, from http://www.humanitarianresponse.info/system/files/documents/files/Comparison%20Tool%20Guidance%202015%20FINAL-EN.pdf.

Zinnert, S., H. Abidi and M. Klumpp (2011). "Data Envelopment Analysis for Humanitarian Logistics." Maritime Logistics in the Global Economy: Current Trends and Approaches 5: 423.

8 July 2015

The demonstration materials are found in *"Acaps_DEA_DemoFiles_150707.zip".*