

Closing the Data Gap: A Model for National-Level Food Loss and Waste Data Generation in the absence of Existing Data

Dr.ir. JM (Han) Soethoudt

January 2024

1. Introduction

In the UN many countries committed themselves to reduce FLW towards 2030¹. So-called national pathways are developed, written by governments, elaborating on how this will be established. In order to support and validate these efforts, FLW monitoring² is required on national level. This is a complex task for LMIC, since it is difficult to get data from rural and remote areas. In addition, the variety in products, supply chain links (SCLs), climatic conditions (weather, water, soil) increases the workload of any monitoring approach dramatically.

Lack of data make it particularly difficult for LMIC, to specify the hotspot food products and chain stages, to define smart targets and to identify adequate interventions. In order to contribute to this essential information we developed a mass flow model based on secondary data to derive the volume of FLW and the associated parameters accordingly. This approach allows to present an indicative country profile showing per food product category and chain stage not only the amount of FLW but also the GHGEs, the land-use and water footprints related to producing the FLW as well as induced nutrient losses. Focusing on food product (sub)categories and chain stages which largely contribute to the aforementioned parameters can substantially lead to resource use efficiency and at the same time to climate mitigation action and nutrition security. Currently, this approach is applied to four countries: Vietnam, China, Kenya and Colombia. Since these country profiles are not on product level, it is not immediately clear, where to start your intervention; it helps to focus by funnelling the options. The step that identifies the actual product(s) that have the highest FLW contribution, requires additional work.

Many quantifications methods are laborious. Surveys, interviews and measurements require huge investments in time and costs. This might be a threshold for many countries, leading to less effort to reduce FLW. In this document we suggest input-output modeling FLW to overcome this issue.

Some relevant considerations are:

- A suggestion is to focus on the top ten products grown/caught in the country. The assumption is that FLW in absolute weight is highly correlated to the national production weight.
- Stakeholder networks exist in various countries, and can be reached by cell phone.
- Lots of data are already collected, in particular by FAO at farm level. Mostly, characteristics of the livelihood (social aspects: age, education, farm size,...)
- Compared to current data, better accuracy is required with respect to FLW data. At least, changes per year should be detectable.
- There are static characteristics (not changing (much) over the year) like (e.g. on farm level) educational level, age, years of experience, having storage, way of harvesting; and there are

¹ <https://www.unfoodsystemshub.org/member-state-dialogue/dialogues-and-pathways/en>

² FLW monitoring is FLW quantification in weight or in % of the incoming flow/production. If we talk about quantification, both indicators are included.

characteristics that (can) change a lot within a year like the weather conditions, price, (competitive) supply.

- Relevant inputs for the models will sometimes be similar, but will surely also vary with the country, product and part of the supply chain. E.g. diseases and distance travelled between SCLs. Note that in case of similar inputs it is likely that the parameter values of the model per country show differences.
- Causes of FLW are often taken as inputs for the model, and there is a clear overlap with the total set of (logical) inputs. An analysis on causes is carried out in the EFFICIENT protocol³, but also in literature. It might be useful to identify the root cause as the input.

If combined with FLW data, understanding the relations between variables and FLW can be examined, and exploited to model FLW. Modeling has the following advantages:

- It can give an understanding what causes contribute the most to FL
- It reduces the amount of work for data collection
- Results from one country can help other countries, since causes may overlap per 'product-supply chain link' combination and (statistically) techniques can be shared and improved to help every country.

2. Modeling

There have been efforts to model FLW with various goals [1-16]. It can be very specific, e.g. for an online retailer in the Netherlands or a group of rice farmers in Nigeria. But that is not the goal here. The FLW needs to be quantified on a national level to meet the SDG 12.3 goals: it calls to 'halve per capita global food waste at the retail and consumer levels and reduce food losses along production and supply chains, including post-harvest losses'

a. Model requirements

There are three dimensions that have specific characteristics in relation to FLW:

- Product
- Country
- Supply chain link

This implies that for each combination a separate model is required. There is no model that describes FL for milk and at the same time for bananas, since the responsible causes/variables/inputs will be totally different. The same holds for farmer vs retailer. Of course, for one country the second dimension is not there and the number of combinations is decreased. Models may differ per country, because of climate, soil, policy, etc., but it is likely that models for countries in the same region show similarities.

Product: in this research a product is considered by its name and not by its variety, brand, way of production or certificate. A banana is a banana, and not a fair trade banana. Note that in the case of FL in LMIC most products are not combinations of ingredients, like frozen pizza or confectionary. The advice is to start with the production top 10 commodities (in tons).

Supply chain link: there are many possible types of stakeholders in a food supply chain, of which a bunch is not mentioned in the mainstream literature on FLW monitoring (e.g. exporter, importer, transport company). The selection of supply chain links in this research is based on the FAO supply chain description:

³ <https://www.wur.nl/en/research-results/research-institutes/food-biobased-research/show-fbr/take-the-target-measure-act-approach-to-reduce-food-waste-yes-but-be-pragmatic-about-it.htm>

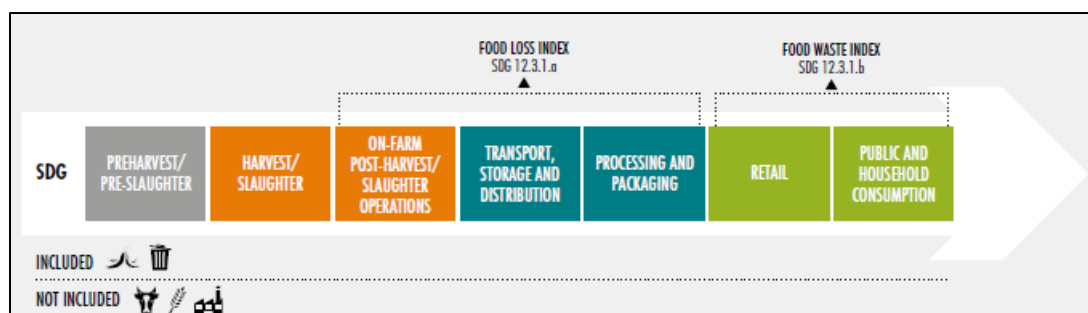


Figure 1: FAO food supply chain description

There are two reasons why this scheme is our starting point. Firstly, it is used by SDG 12.3, which is the reference for the FL and FW targets in 2030 and corresponding national pathways. Secondly, it relates to FAO data collection. FAO has done some modeling on FLW. They could do so, because they have lots of data of LMIC all over the world, especially at farm level. There is no other entity collecting data of this kind worldwide, and hence their structure is a good *starting point* for FLW modeling. However, later on when performances of models become clear some changes could be necessary for data collection. Bottom line is that many LMIC countries don't have resources to set up data collection at the required level of detail. In that context, it is important that FAO shares the national data with the country at hand, which is a necessary condition for a kick start. We believe that if these data are not available no target will be reached, since the data collection is still not in place for whatever monitoring you want.

Looking at [Figure 1](#) we see that the supply chain is described by activities rather than stakeholders. We link them as follows:

Table 1: linkage between food supply chain activities (food loss part) and stakeholders

Food supply chain activities	Stakeholder
Harvest/slaughter + on-farm post-harvest/slaughter operations	Producer/farmer/fisher
Transport, storage and distribution	Intermediary/trader
Processing and packaging	Processor

Although other types of food supply chains (FSC) exist, the one in [Figure 1](#) is the most common one, and will be used for national FL monitoring of LMIC, by modeling the FL at the stakeholders in [Table 1](#).

b. Technical part:

Based on some limited literature review, we conclude that the research on FL and/or FW modeling is scarce. The results so far don't show a breakthrough, in the sense that it is an accepted alternative for other quantification methods. In our opinion, the reason for that is, that the performance of these models is not proven yet (we couldn't find any result with respect to the accuracy of the FLW model quantification).

Various types of models can be found in relation to FLW. There are linear input-output models to estimate the FLW based on a number of inputs. The performance of these models is very unclear (especially in terms of explained variance). Another model that is applied to FLW is Probit-regression. This is used to identify factors that have a strong relationship with FLW. A rather new approach is the Interpretive Structural Modeling methodology, also applied to identify causes and root causes for FLW, based on expert opinions. Also more theoretical models are tested (see references).

In the CGIAR project we propose finding estimates of FLW. This suggests a quantitative input-output model where the output is a number, and not the name of a (root) cause or a determinant. Obviously, finding causes and determinants for FLW is very useful, since they are potential input variables for the input-output model.

The FAO (and IFPRI also a little) is working on linear multivariable input-output models (linear regression) in the context of the Food Loss Index [16]. They try to help countries with monitoring FLW, by improving their FAO models. Going from global/regional to country level data is their way forward. This requires more data on country level (to explain differences between countries in the same region).

The problem is that one needs many datapoints over the years, to fit and test the model. Otherwise the model is over dimensioned. These data are not available yet, since collecting data on country level just started in LMIC (by the FAO). The FAO is solving this partly by introducing time-invariant dummy variables. They also included a list of input variable categories:

- Energy
- Inputs (fertilizer, etc.) and other associated costs
- Investment and monetary policy
- Social and economic factors
- Storage
- Transportation and logistics

Note that these variables are related to stakeholders in the entire supply chain. Their models try to estimate the national number of FLW in total, i.e. one number for the total FLW of all products and supply chain links. It has to be investigated if the degrees of freedom are less here, because we propose to model per product-country-SCL combination.

Another technical issue is the difference between static and time-variant inputs. Education level is static, but average temperature is time-varying over and within the year. For some inputs the variations within the year should be translated in a value for a year, since that is the level of detail in time.

Based on the above our suggested modeling approach looks as follows:

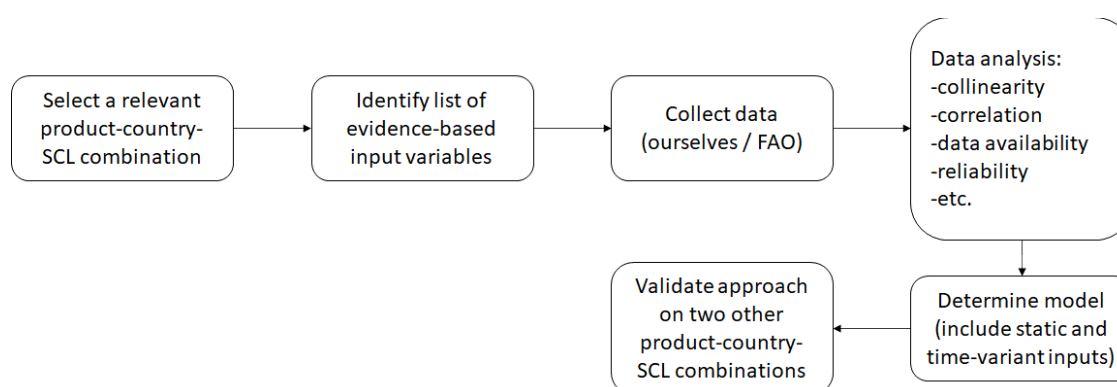


Figure 2: approach for the model

The first step can be derived from the country profiles, as described in the Introduction.

3. Data

The goal is to apply (in first instance simple linear types of) models to understand relations between inputs and FLW per combination of country-product-supply chain link. If the explained variance is high, sampling input data can be used to get insight on FLW on national level, which is relatively easy (and partly static) compared to collecting FLW data.

Output data (FLW)

To build input-output models a statistical analysis is required. Reliability often comes with the amount and quality of available data. With respect to the output data of the model (i.e. the FLW) real measurements are scarce (since they take time and are costly), and in most cases estimates are provided by stakeholders and/or experts. Although FLW data collected by interviews and surveys are less reliable (and usually underestimates), it is a necessary concession when modeling input-output relations of the target population of stakeholders is large (e.g. 1,000,000 farmers) or/and FL collection is not organized nor centralized (e.g. for large groups of consumers/inhabitants in LMIC countries). Nevertheless, in a relative context, we believe the right priorities for FLW reduction interventions will emerge.

Input data

As described earlier there are various types of input data that possibly relate to the amount of FLW. If the data are static, they have to be collected once (and again after e.g. 5 years). Inputs that vary per year or at shorter notice are gathered more frequently.

4. References

1. Zhang, R.J., B. Lee, and H.-H. Chang, *What Is Missing in Food Loss and Waste Analyses? A Close Look at Fruit and Vegetable Wholesale Markets*. 2019: p. 13.
2. Chalak, A., C. Abou-Daher, and M.G. Abiad, *Generation of food waste in the hospitality and food retail and wholesale sectors: lessons from developed economies*. 2018: p. 12.
3. Samadder, A.K.S.R., *An empirical model for prediction of household solid waste generation rate – A case study of Dhanbad, India*. 2017: p. 13.
4. Gokarn, S.K., T. S., *Analysis of challenges inhibiting the reduction of waste in food supply chain*. *Journal of Cleaner Production*, 2017. **168**: p. 595-604.
5. GL, S., et al., *Farmers' Perceptions of Rice Postharvest Losses in Eastern Uganda*. 2017: p. 14.
6. Berkenkamp, J. and C. Phillips, *Modeling the potential to increase food rescue: Denver, New York city and Nashville*. 2017: p. 41.
7. Delgado, L., M. Schuster, and M. Torero, *Reality of Food Losses: A New Measurement Methodology*. 2017: p. 49.
8. Delgado, L., M. Schuster, and M. Torero, *Quantity and quality food losses across the value Chain: A Comparative analysis*. 2021: p. 16.
9. Delgado, L., M. Schuster, and M. Torero, *On the origins of food loss*. 2021: p. 31.
10. Shee, A., et al., *Determinants of postharvest losses along smallholder producers - maize and Sweetpotato value chains: an ordered Probit analysis*. 2019: p. 20.
11. Magalhaes, V.S.M., L.M.D.F. Ferreira, and C. Silva, *Using a methodological approach to model causes of food loss and waste in fruit and vegetable supply chains*. 2021: p. 13.
12. Meijboom, J., *Waste reduction in e-groceries fulfilment centers: a case study at Picnic*. 2019. p. 152.
13. Kojima, R., et al., *Theoretical model of food loss and waste (FLW) generation in household and empirical evidence*. 2021. p. 34.
14. Mingion, M., C. Fabi, and G.J. Lasinio, *Measuring and Modeling Food Losses*. 2021: p. 42.
15. Carlsson, A., *Predictive Regression Model Evaluation - Evaluating Predictive Machine Learning Models to Reduce Food Waste in the Dairy Industry*. 2023. p. 88.
16. Taglioni, C., J.R. Moncayo, and C. Fabi, *Food loss Estiation: SDG 12.3.1a Data and modeling approach*. 2023. p. 52.