2021

GIS MCDA - Tanzania

Dairy processing industry – V.1.1

NELSON RIBEIRO

Contents

Figures index

Tables index

Introduction

This report documents a Geographical Information Systems - Multicriteria Decision Analysis (GIS-MCDA) for the identification and definition of dairy processing industry (UHT and milk powder) potential sites in Tanzania.

Proposed research question can be formulated as:

What would be the best site to locate a dairy processing unit (UHT and milk powder) in Tanzania?

Livestock and dairy sector context and background are briefly detailed based on literature review. Following previous exercise for the Sub-Saharan Africa Mobile Storage Location Project (Ghana, Kenya and Tanzania) the modeling assumes a spatial or GIS-MCDA methodology using weighted factors (AscoughII et al., 2019; Boroushaki & Malczewski, 2010; Malczewski, 2006). Introduces dairy sector business rules and minor changes in top score areas selection method and adding new final exclusive criteria.

Data modeling variables characterize supply, demand, and infrastructure/accessibility by defining the following criteria:

1. Supply
    a. Feed - crop production aggregate (Wood-Sichra et al., 2016)– crop byproducts and forage production areas.
    b. Livestock production systems (Robinson et al., 2011) – potential intensification areas.
    c. Dairy herd spatial distribution – estimates on animal distribution (Cattle) GLW 2.0, Cow number: (Tanzania National Bureau of Statistics, 2017).
2. Demand - Human population density, large cities, urban areas.
3. Infrastructure - Transportation network (accessibility).

The transportation network infrastructure is modeled as raster-based travel time/cost analysis (Mulrooney et al., 2017).

Accessibility/infrastructure travel time/cost to market is processed for large national cities. Access to finance - distance to bank agencies – and major road proximity are complemented by access to IT access - mobile broadband coverage – and urban areas to filter location score grids into recommend areas of interest.

In a final exercise step to reduce areas of interest into single sites, large-scale visual map and satellite imagery inspection is conducted (Google, Bing and OpenStreetmap), to pinpoint optimal locations. This value judgement assessment uses as criteria: (i) port and railway station proximity; (ii) logistics and industrial districts/areas; (iii) availability of space. This can be understood as a returning from modeling abstraction or reduction to field reality and a first step into operationalization.

To understand the model consistency and explain outputs, the final results are overlayed with existing dairy industrial facilities.

The project is developed using open-source GIS software QGIS 3.10.12-A Coruña with open-data sources OpenStreetMap and FAO GIS platform, mobile broad band coverage data is derived from Collins Bartholomew's Mobile Coverage Explorer datasets.

This document is structured in 4 main sections: 1. Initial Specification; 2. Data Sources; 3. Vector Data Preprocessing/Editing; 4. Data Editing Geoprocessing, with Introduction and Closing Remarks. In Annexes there are: project flow diagram, algorithm diagrams and complementary map outputs.

GIS Multicriteria Decision Analysis

Spatial decision problems involve a set of geographically defined alternatives and multiple and sometimes opposing assessment criteria. Alternatives are commonly assessed by many intervenient (decision-makers, stakeholders, interest groups).

GIS multicriteria decision analysis GIS-MCDA consists of a method to convert and combine spatial data/geographical information and decision-makers criteria to attain evidence for a decision-making process. GIS capabilities are enhanced by MCDA procedures, techniques, and algorithms for structuring decision problems, to design, evaluate and prioritize alternatives.

Integration of GIS and MCDA provides a replicable model, improves communication between project participants or decision-makers, can offer a different perspective of problem and solution, helping to redefine initial specification and/or criteria.

GIS multicriteria analysis methods are usually presented in a three-stage hierarchy of: intelligence, design, and choice.

In the intelligence phase, data are acquired, processed, and exploratory data analysis is performed.

The design phase should entail the formal modeling/GIS interaction development of a solution set of spatial decision alternatives. The integration of decision analytical procedures and GIS functions is critical for supporting the design phase.

The choice phase involves selecting location alternatives from those available. Specific decision rules are used to evaluate and rank alternatives.

The three stages of decision making do not necessarily follow a linear path from intelligence, to design, and to choice.

In the case of the present exercise we would place it in the first stage of this hierarchy, the exploratory/intelligence phase.

From a critical standpoint it can be stated that, while quantitative data analysis and evidence gathering through GIS modeling certainly contributes to attaining evidence for decision-making processes, in reality, a complex set of socio-economic, political, cultural, ethno-anthropological aspects and power relations actually shape processes and govern decision-making.

Modeling is also as good as the input data, its quality and reliability support the extent to which conclusions can be trusted, and these are just as sound as the analysis conducted. From that prism, specification and objectives define the modeling assumptions and approximations and can always produce distinct answers (Kitchin, 2014b).

Data are both social and material do not just represent the world but can actively produce it, are not mere raw material of information and knowledge, do not exist independently of ideas, techniques, technologies, people and contexts that produce, process, manage, analyze and store it. Positionality is always present even when "data speaks for itself" (Kitchin, 2014a).

1. Tanzania livestock and dairy sector context

Agriculture is the main economic sector in Tanzania representing 29% of the GDP in 2018[1] generating 70% of the total income for rural households. In 2017 Around 4.6 million households, 62% rural and 23% urban, owned livestock, 35% of which owned cattle (ILRI et al., 2017). The livestock sector employed 50% of the population with a 30.5 million cattle heads population, about 1.4 % of the global and 11% of the African cattle population.

Still, milk consumption is low compared to neighboring countries and according to WHO guidelines, most milk traded through the informal market and only 2.7% of the raw milk processed. Production is essentially destined for household consumption the marketed milk is sold primarily directly to neighbors and small-scale informal milk traders/hawkers. Processors run at very low levels of utilization – often as low as 30% (vs. 70% best practice utilization levels for manufacturing assets) (Dalberg et al., 2019), with the country importing about half of dairy processed products, mostly from Kenya.
A large portion of milk marketed comes from small-scale livestock farmers (70%) and development is limited by poor nutrition and support services, and the insufficient supply of dairy stocks. Until 2015 despite contributing with 30% of the annual milk production, improved dairy breeds corresponded to about only 3% of the national cattle herd. (Lunogelo et al., 2020). Other constraints include inadequate financial and credit services and facilities, poorly organized milk collection and distribution, and high incidence of livestock diseases.

It's not surprising that a literature review shows a wide consensus on the existence of a huge demand potential for dairy processed products and a high return on investment in the sector (Dalberg et al., 2019; ILRI & CGIAR, 2017; Kurwijila et al., 2012; Nell et al., 2014; NIRAS, 2010; Njombe, A.P., Yakobo Msanga Nathaniel Mbwambo, 2011; Quaedackers et al., 2009). Several

[1] CountryProfile | World Development Indicators (worldbank.org) https://databank.worldbank.org/views/reports/reportwidget.aspx?Report_Name=CountryProfile&Id=b450fd57&tbar=y&dd=y&inf=n&zm=n&country=TZA

factors contribute to this understanding. First, the low *per capita* consumption, second there is a fast growing urban population and middle class, third, a large margin to change consumption from raw to processed milk, fourth and finally, there are clearly identified regional export markets, the DR of Congo, Malawi and Mozambique.
There is also a general agreement that an increase in investment is needed to boost productivity and income for producers, and to enhance the sector economic contribution to development objectives (poverty; food security, economic growth, exports, industrialization and employment).

The Tanzania Livestock Sector Analysis (TLSA) conducted a herd and livestock modeling and assessment for 2016 and 2017 and the long-term potential for its development (ILRI et al., 2017). According to the assessment, that informs the Tanzania Livestock Master Plan (TLMP) (ILRI & CGIAR, 2017), a business as usual scenario projects a large supply deficit in milk, driven essentially by poor genetics, health and feed constraints, resulting in a total production-consumption gap/deficit of 5.8 million liters in 15 years.

To counterbalance those projections and stimulate the sector development the TLMP Dairy Development Roadmap 2017/2018–2021/2022 defines an overall vision of an increased milk production that meets domestic demand and generates a surplus. This vision could be attained by increasing dairy cow productivity through improvements in genetics, health and nutrition, expanding the national dairy cow herd and improving the processing and marketing of dairy products.
It sets ambitious goals for the period: an increase of crossbred dairy cattle by about 3.8 times from 783,000 to 2,985,000, a boost in production from 2,159 million litres to 3,816 million litres - about 77% - and an growth in the annualized milk production of a cow from 179 litres to 254 litres.

The Plan proposes transforming traditional family dairy farms into improved market-oriented family dairy systems, modernizing the sector with interventions in the areas of feed, genetics, health, promotion, marketing and policy. The future scenario analysis with investments projects closing milk consumption requirement and generating a surplus, based on a list of proposed interventions, artificial insemination (AI), synchronization, multiple ovulation and embryo transfer combined with improved feed and health interventions, more investment in value addition and complementary policy changes.

Among the policy priorities define in the TLSA is the creation of a conducive environment for investment in commercial milk production and processing. The TLMP assumes it as a critical condition for the plan success: encouraging the private sector to invest in milk processing plants and dairy farms and, the promotion of the investment in long shelf-life milk products like UHT and powdered milk. This investment need is justified by the short shelf-life product availability and fluctuations in supply due to seasonality. The total investment to improve milk marketing and processing was budgeted in over 45 million USD.

1.1 Dairy Geography

The TLSA classifies production systems using Seré and Steinfeld (Robinson et al., 2011) approach into three livestock production zones, namely, Central, Coastal &Lake, and Highlands. In addition, there is a commercial/specialized livestock production system, the urban and peri-urban feedlot production system, across all the production zones.

Most farmers are pastoralists and only in highlands, where crossbred cattle are present and milk production is higher, the zero-grazing principle is applied. The TLMP breeding intervention targets Highland and Coastal&Lake zones only. Harsh climate conditions, limited availability of crop residues to support dairy crossbreeds make intensification in the Central zone not plausible. In both Highland and Coastal&Lake, it is assumed that 90% of the additional feeds will be improved forages (grass/legumes/fodder trees and shrubs) and the remaining 10% locally formulated concentrates and industrial by-products.

Investments in breed and feed improvement in Coastal&Lake and Highlands are framed by the promotion of collection of milk cooperatives and associations and, the investment in milk processing plants.

In summary geography impose genetic, feed, and health conditions that lead to the assumption that dairying intensification is more likely to occur in the Lake&Coastal and Highlands production zones. Primary production zones permit crossbred dairy cows, a crop production that can supply feed byproducts, and the possibility of growing forage crops, summing to this, thanks to a lower presence of vectors and the presence of production systems in the zero-grazing principle, there is a much lower incidence of cattle diseases and zoonoses.

Nonetheless, commercial/specialized urban and peri-urban feedlot production systems can occur in all zones, with the limiting factors being accessibility to input and output markets.

Milk processing is also dependent on a stable and reliable milk supply to run at recommended levels of utilization to be able to compete with imports and informal trade. Milk collection and chilling centers role in the value chain is significantly susceptible to economic, ethnoanthropological and socio-cultural factors like pricing, trust, ownership, and power relations (Quaedackers et al., 2009).

Finally, some caution and debate should be considered around the impact of production intensification and industrialization. Increasing demand based on the fast growing urban and middle-class population typically imposes a change from extensive, small-scale, subsistence, mixed production systems to intensive, large-scale, specialized production units. Sector industrialization leads to vertical and horizontal concentration, attracts large multinational firms, and changes production spatial patterns.

This transformation, while increasing production, can have adverse nonplanned consequences. First, it can have an extremely negative effect on livelihoods. This can be particularly severe for the small-scale/smallholder family-based production fabric, when considering that around 1.6 million households own and depend on cattle and, with a very high poverty incidence (39% rural households living below ‘basic needs’ poverty line in 2011(Pica-Ciamarra et al., 2011)).

Second, it can also dramatically increase the sector negative environmental impacts, from both, the pressure on input resources and the output pollution effluents and GHG (Food and Agriculture Organization of the United Nations (FAO) & New Zeland Agricultural Greenhouse Gas Research Centre, 2019; Robinson et al., 2011).

2. Initial Specification

The general data dimensions specified for the mobile storage analysis were the following:

1. Infrastructure dimension
   a. average lights
   b. primary roads
   c. railways
   d. airports
   e. bridges
   f. waterways
2. IT access
   a. internet access
   b. cellphone signal
   c. broadband information
3. Access to finance
   a. bank locations
4. Market access
   a. cities and travelling time to cities
   b. time to ports
   c. port waiting times
5. Population density
   a. Socioeconomic information maybe included if available at subnational level (poverty maps for example).
6. Production dimension
   a. Cereals and vegetables production
   b. Crop calendars and plant phenology.
   c. Livestock

3. Data Sources

The following sources are used in the modeling:

1. ***Human Population Density 2020*** – WorldPop2020 - Estimated total number of people per grid-cell 1km. https://www.worldpop.org/geodata/summary?id=24777.
2. ***Mapspam Production*** – IFPRI's Spatial Production Allocation Model (SPAM) estimates of crop distribution within disaggregated units (Wood-Sichra et al., 2016). https://dataverse.harvard.edu/dataverse/harvestchoice
    - Production (mt).
3. **GLW Gridded Livestock of the World** - Gridded Livestock of the World – Latest – 2010 (GLW 3) (harvard.edu) (Gilbert et al., 2018).
4. **Global Livestock Production Systems v.5 2011** (Robinson et al., 2011).
5. ***OpenStreetMap*** **-** Map of the world built by volunteers and released with open-content license. Community mapping using wiki-style collaborative editing software. The data was downloaded in ESRI SHP file format for the selected countries available at: http://download.geofabrik.de/. (Ramm, 2019).
6. ***FAO*** – FAOStat, Rivers of Africa, Inland Waters of Africa, Airports, Ports.
7. ***Collins Bartholomew*** - Mobile Coverage Explorer raster data representation of the area covered by mobile cellular networks around the world. The dataset series is supplied as raster Data_MCE (operators) and Data_OCI (OpenCellID database).

4. Vector Data Preprocessing/Editing

This step is partially automated and includes data gathering/download, selection and edition by location and attribute, and the creation of a country database *vector geopakage*[2]:

1. ***OSM Road layer (gis_osm_roads_free_1.shp)*** – Selected by location and attribute to generate a major roads layer. A comprehensive description of the features can be found in (Ramm, 2019). Lack of data on road network conservation, quality and speed limit for most of the network imposes a conservative approach.

    1.1 Attributes - 'motorway'; 'trunk'; 'primary', 'secondary'.

2. ***OSM Railways (gis_osm_railways_free_1.shp)*** – Selected by location.

3. ***OSM Point-of-Interest Layer (gis_osm_pois_a_free_1.shp)*** – Select by attributes to generate layer Banks.

    3.1 Attributes: 'bank'.

4. ***OSM Transport Layer (gis_osm_transport_free_1.shp)***

    3.1 'ferry_terminal'; 'railway_station'; 'railway_halt'.

5. ***OSM Places Layer (gis_osm_places_free_1.shp)*** - Selected by attribute to generate major human settlements layer.

    5.1 Attributes: 'city'; 'town'; 'national_capital'.

6. ***FAO Data - Ports; Airports; Secondary Airports*** - csv file formats - FAO http://rkp.review.fao.org/geonetwork – selected by location for the country.

7. ***FAO Major rivers*** - Rivers of Africa derived from the World Wildlife Fund's (WWF).

[2] http://www.geopackage.org/

8. ***FAO Inland Waters*** – Clipped by country bondaries.

Data is edited extracted/clipped using country borders.

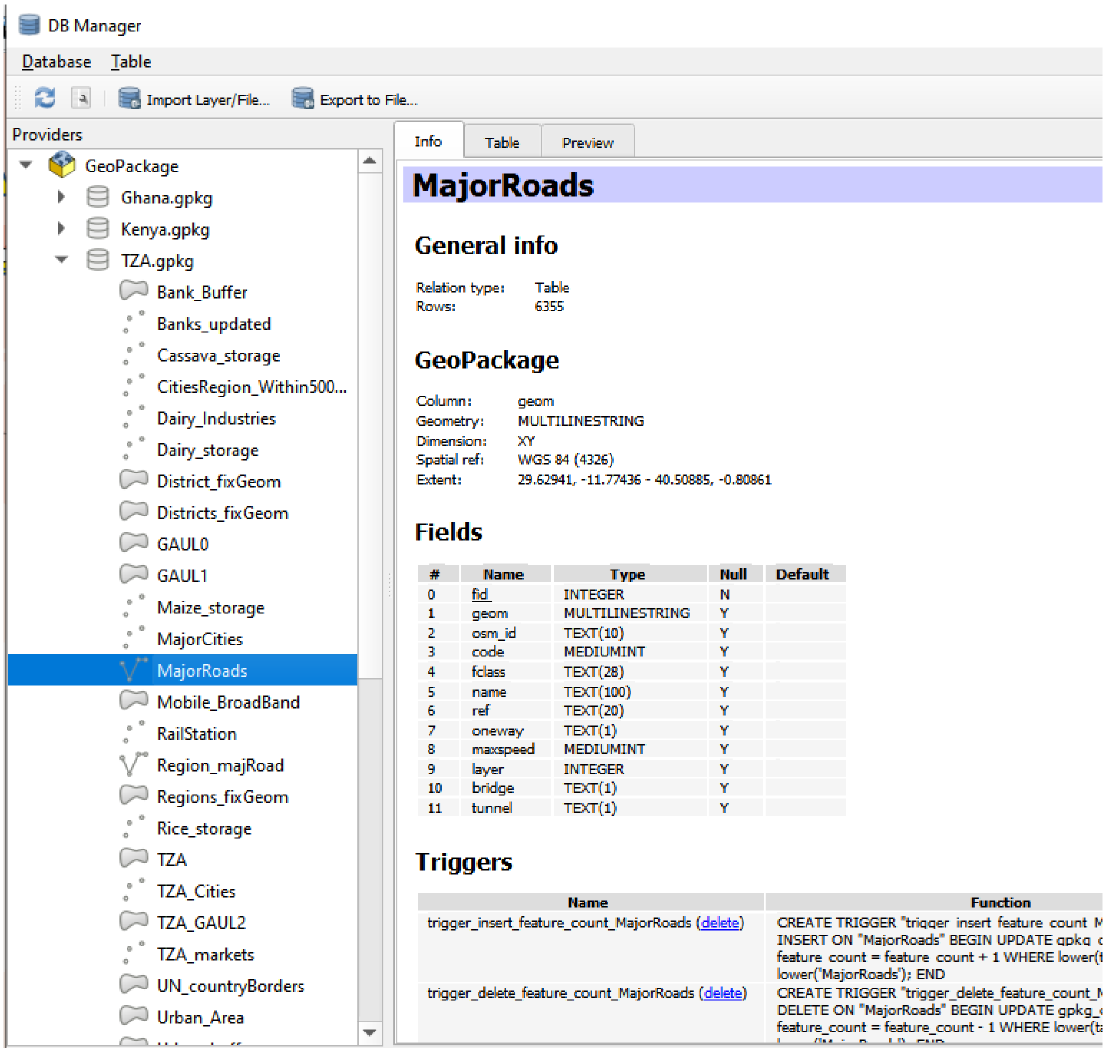

Figure 1 - Geopackage

5. Data Editing Geoprocessing

This section details the editing and geoprocessing steps.

5.1 Raster travel time surfaces – Cost/time distance calculation

The development of an algorithm (python script) partially automates the accessibility map data processing (travel time/cost surfaces).

The calculation of time/cost distance surfaces is based on some assumptions:

1. Cities are points features, lakes (inland waters) represented by polygons, other infrastructure network layers consist of linear features. Alternatively, railway stations (points) could also be used as railway connectivity imposes a tunnel effect, time-distance compression between points (stations), with low accessibly in between.
2. Major river navigation is not computed.
3. Road travel time/cost is modeled for trucks, it is assumed that the accessibility is related to cargo freight vehicles, tertiary and local traffic roads are not included.
4. Lake and river navigation are treated as surface (polygons) not taking into consideration navigation infrastructure (points), it is assumed for small to medium cargo crafts.
5. Major islands accessibility is computed, ocean navigation input parameter value is the same as river and lake navigation.

The general steps to produce accessibility maps (travel time surfaces) are rasterization vector layers, creation of cost friction surface, and computation of a cumulative time/cost layer from/to points:

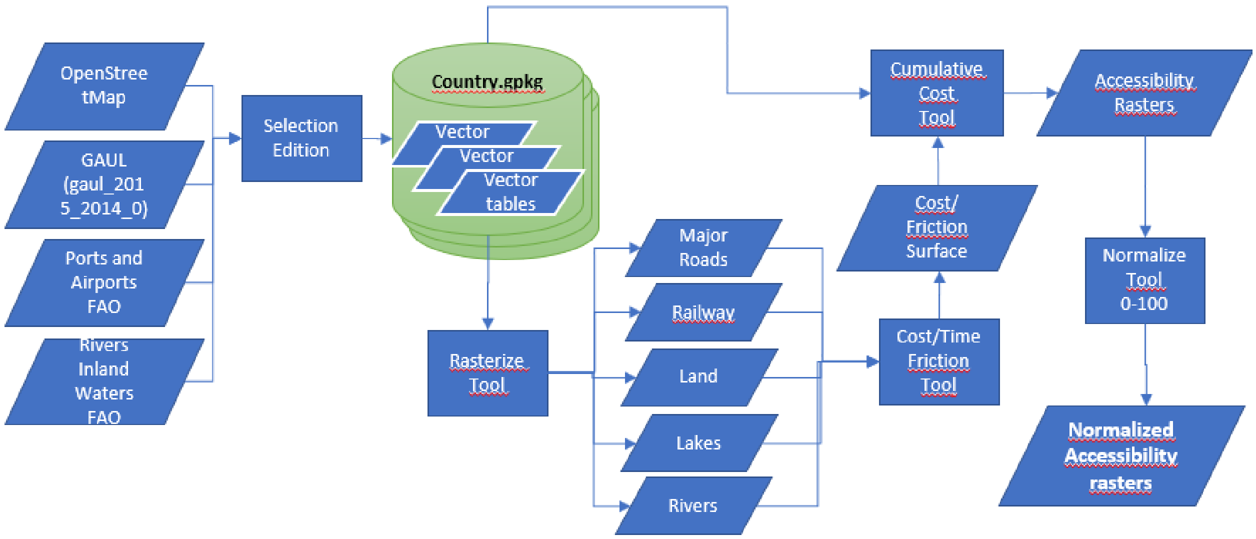

Figure 2 - Accessibility modeling flowchart

1. ***Rasterize* Tool** *(SAGA raster normalization tool)* – Communication/transportation network and surfaces conversion from vector to raster with 1km cell grid burning a value of average time (minutes) to travel 1km for the considered transportation mode ((a) land/walk, (b) major roads/vehicle, (c) railway/train, (d) navigation.

   The assumed modeling values:

Land (a)	10
Major roads (b)	1
Railway (c)	0.6
Navigation (d)	3

The rasterization outputs individual 1km raster grids with the modelling value per individual cell. Modelling value - speed - parameter can be easily changed/adapted to a different specification.

2. ***Cost/Friction Tool*** *(GRASS r.series tool)* – A cost or friction surface or grid is obtained overlaying (a), (b), (c), (d) grids, propagating the minimum cell values; The cost/friction surface output covers Tanzania and bordering countries.

3. ***Cumulative Cost Tool (****GRASS r.cost tool)*– Service Area – The cumulative cost/accessibility maps are produced selecting a central point, or points, and defining service areas, outputting an accessibility grid for Tanzania major cities[3] (>200k habitants).

City	Population	Weight %
Dar es Salaam	4,364,541	0.579
Mwanza	706,453	0.094
Zanzibar	501,459	0.067
Arusha	416,442	0.055
Mbeya	385,279	0.051
Morogoro	305,840	0.041
Tanga	221,127	0.029
Kigoma	215,458	0.029
Dodoma	213,636	0.028
Songea	203,309	0.027

Table 1 - Population major cities Tanzania

Cross border trade (demand), accessibility to large regional cities is not considered since Kenya, Burundi Uganda and Rwanda have strong dairy sectors. Congo DR, Malawi, and Mozambique can become import partners in the medium to long term.

4. ***Normalization*** – Different criteria units are normalized scaling from 0 to 100 to calculate the location score (weighted sum). The lowest accessibility value (time or cost) corresponds to 100 (high accessibility).

[3] https://www.citypopulation.de

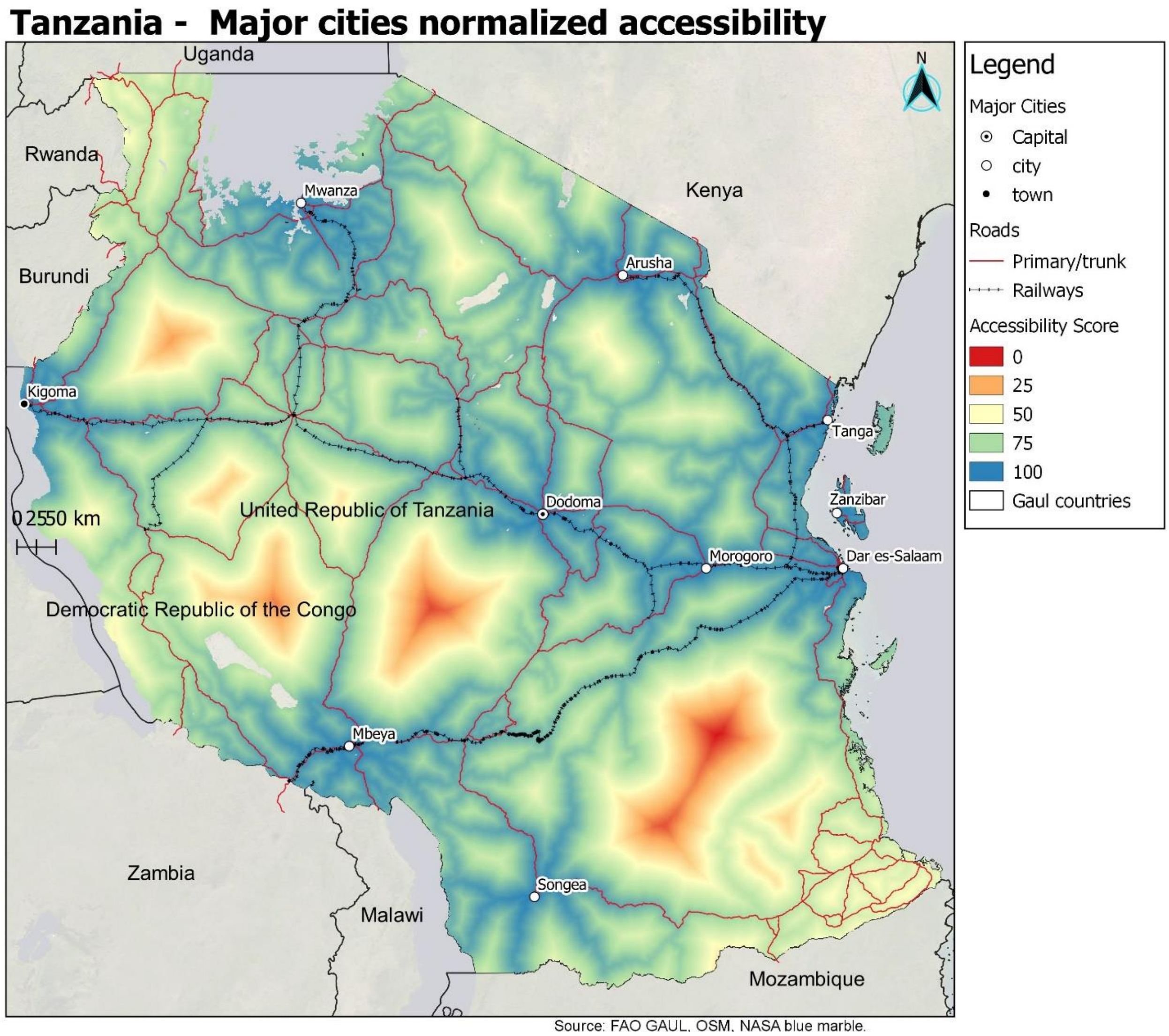

Figure 3 – Normalized accessibility cities

5.2 Human Population Density

World human population density estimates for 2020 1km grid raster.

1. ***Editing*** – Clipped/masked using GAUL country borders to extract country grid.
2. ***Normalization (raster normalization)*** – Raster grids are normalized into common scale (0 to 100) to calculate a multicriteria location score (weighted sum).

In the case of population density, a high concentration of population in major cities, specially Dar es Salam, implies a frequency distribution extremely left skewed, a natural break/natural classes method of classification can allow an informed decision (expert value judgement) on thresholds of urban rural definition.

5.3 Production - Dairy

The analysis focus on areas with intensification potential for the location of large dairy processing units for UHT or milk powder.

Potential intensification areas are defined using crop, production systems, dairy herd density.

- MapSpam crop data (aggregated production) is used to identify productive agricultural areas.
- Global livestock production systems are select for areas with high LGP and where crossbred cattle can thrive, temperate, humid and highlands (Costal&Lake and Highland production zones).
- Density of dairy herd based on:
  - Cattle distribution: GLW 2.0, 1km res.
  - Cows number - (Tanzania National Bureau of Statistics, 2017).
  - Administrative Divisions: (GAUL_Level1).

The recent construction of new processing facilities and the current capacity underutilization can question the analysis relevance. While the TLMP investment plans should have come to boost processing capacity utilization through increasing productivity and milk traded trough formal channels, it also promoted the increase of the processing capacity and the number of units, with 15 new processing plants set up in 2018/2019 financial year[4]

Existing dairy processing units were georeferenced using

a) Eastern and Southern Africa Dairy Association (ESADA) Dairy Africa portal, Tanzania Dairy Processing Status for 2018/19[5].

b) Tanzania Livestock Master Plan (Table 69: Milk processing Industries in Tanzania (2016/2017)).

c) Google, OSM and Bing maps.

[4] https://www.ippmedia.com/en/news/dairy-sector-uplift-15milk-processing-plants-set

[5] https://dairyafrica.com/

d) Tanzania NBS administrative boundaries.

AEZ/administrative production zones were produced using

a) NBS data,
b) Mogriculture[6] and
c) citypopulation.de[7].

5.3.1 Global Livestock Production Systems

FAO Global Livestock Production Systems (Robinson et al., 2011) uses a livestock oriented classification developed by Seré and Steinfeld (FAO 1996) and is the methodology for input data and also used in to classify production systems in the Tanzania Livestock Sector Analysis.

The Seré and Steinfeld (FAO, 1996) scheme classifies livestock systems into four types:

1) landless livestock production systems (LL, which may be monogastric or ruminant).
2) grassland based system (LG,in which crop-based agriculture is minimal);
3) mixed rainfed systems (MR, mostly rainfed cropping combined with livestock).
4) mixed irrigated systems (MI, in which a significant proportion of cropping uses irrigation and is interspersed with livestock).

The livestock-only and mixed farming systems are further characterized in agroclimatic terms, based on temperature and Length of Growing Period (number of days per year during which crop growth is possible). The agroclimatic categories were defined as:

- Arid and semi-arid: LGP ≤ 180 days.
- Humid and sub-humid: LGP > 180 days.
- Tropical highlands or temperate. Temperate: +1 month daily mean temp<5 ºC. Tropical highlands daily mean during the growing period of 5–20 ºC.

The classification system includes eleven types: livestock only, grassland based (LG), which may be arid/semi-arid (LGA), humid/sub-humid (LGH), or tropical high- land/temperate (LGT); landless monogastric-based (LLM), and landless ruminant-based (LLR); mixed, rainfed systems

[6] https://www.mogriculture.com/2016/05/the-role-of-agro-ecological-zones-in-agriculture/
[7] https://www.citypopulation.de/en/tanzania/

(MR) by the three agro-ecological zones, and mixed, irrigated systems (MI), also by the three agro-ecological zones.

Seré and Steinfeld (FAO, 1996) system cannot be mapped directly. First, the classification occurs essentially at the farm level. Second, definitions used in the classification include unavailable data like: 'the amount of farm-produced dry matter fed to animals'. The dataset classification is based on landcover, human population density, length of growing period and elevation data.

5.3.2 Production data processing

1. ***Editing*** – Clipped/masked datasets using GAUL country borders.
2. ***Geoprocessing*** – A dairy herd intensification layer is created using dairy cattle density and the selected livestock production system (LPS used as 1/0 layer).
   A. Selected Tanzania Livestock Production System (TZA_LPS) = (7: MR Mixed Rainfed Humid) + (8: MR Mixed Rainfed Temperate) + (9: MI Mixed Irrigated Hyperarid) + (10 = MI Mixed Irrigated Arid) + (11: MI Mixed Irrigated Humid) + (12: MI Mixed Irrigated Temperate) + (13: Urban).
   B. Raster calculator expression:

   *(("TZA_LPS " >= 7) AND ("TZA_LPS" != 9) AND ("TZA_LPS" <= 13)) * "TZA_dairyHerd" = TZA_dairyHerd_Intensification*
3. ***Normalization (raster normalization)*** – Raster grids are normalized into a 0 to 100 common scale to calculate location score (weighted sum).

5.4 Location Score / Multicriteria weighted sum

The location score is obtained by way of a simple arithmetic weighted sum (*GRASS r.series* tool) of the normalized/scaled grids, location score can theoretically vary from 0 to 100.

The assumed weighting for each of the criteria is as follows.

*("Crop Production" * 0.3) + ("Human Population Density" * 0.1) + (“Major Cities Accessibility” * 0.1) + (”dairyIntensification” * 0.5)*

The final output is an approximately 1km (0.01 degree) raster grid with the location score cell value. Demand factors are low weighted because exclusive criteria use buffers around urban areas.

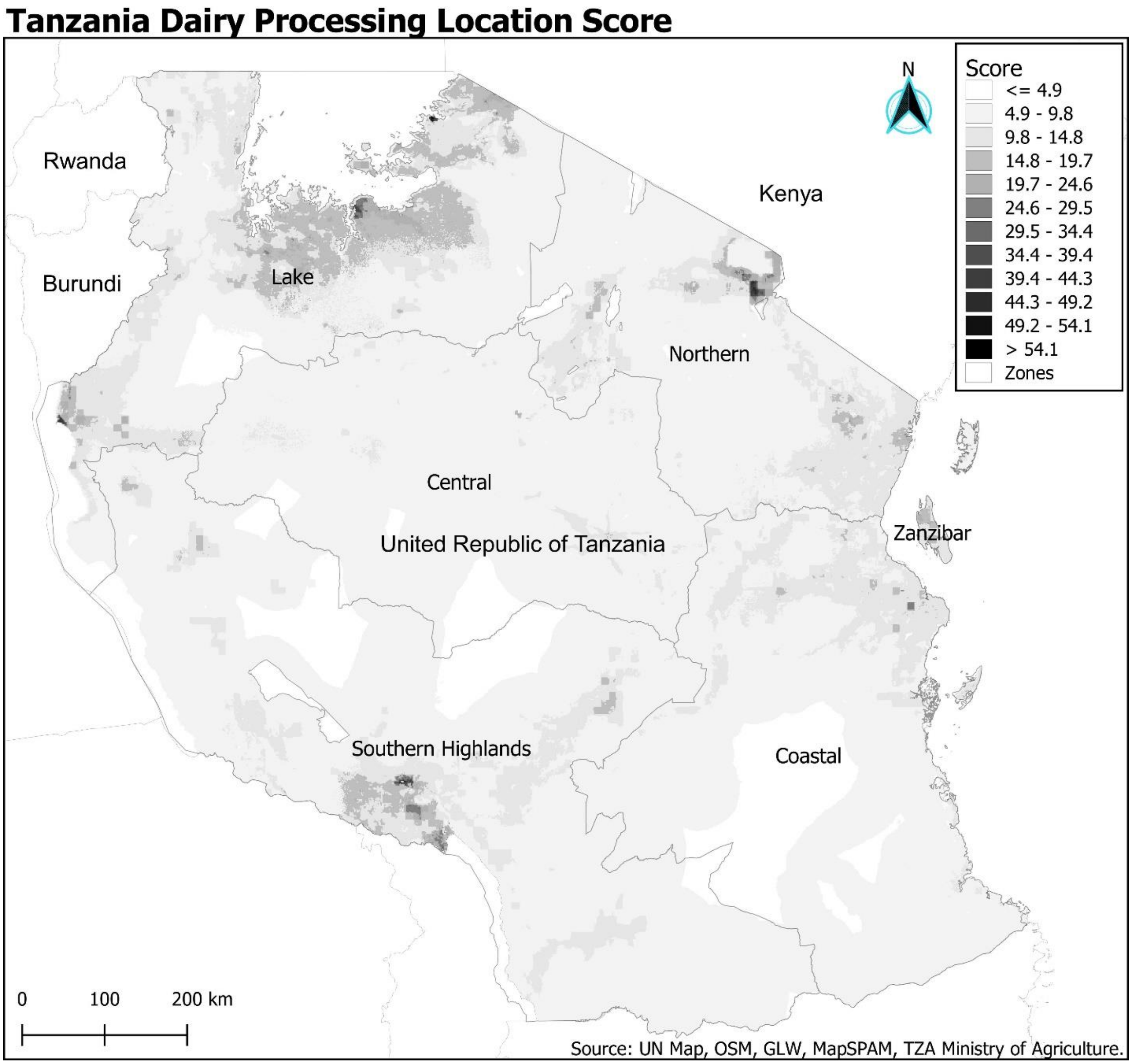

Figure 4 - Dairy location score

Dairy processing location score as a measure of intensification potential clearly identifies Lake and Highlands larger high score areas, while Coastal Tanga, Dar es Salam, Zanzibar benefit from market (demand) proximity and high accessibility.

5.5 Final Location Top Score Map

Final location top score map methodology was further developed from the mobile storage location analysis for Kenya and Ghana. Top score grids are defined using the top 99th percentile and the final location map classifies it further into 3 equal intervals classes. Final locations are selected using the top third class and adding the exclusive criteria: access to finance, distance from major roads, access to IT (mobile broadband connection) and urban areas. Expression: *10Km buffer urban areas AND 20km distance from bank AND 2km distance from roads AND Mobile broadband coverage*

5.5.1 Access to finance and distance to major roads

Access to finance is defined as linear distance to bank agency.

- *Buffering* – the tool computes a buffer area for all the features in an input layer using fixed or dynamic distance:
    - Banks - 20km (0.18 degree) buffer radius.
    - Major roads - 2km (0.018 degree) buffer radius.
- *Intersection* - extracts the overlapping portions of features in the Input and Overlay layers: Banks_Buffer and Roads_Buffer.
- *Dissolve* - Takes the intersection vector layer and combines the features into a new feature, a single polygon.
- *Clip Raster by Mask Layer* – The grids are extracted using the polygon (dissolved intersection of the banks and roads buffers).

5.5.2 Access to IT

Collins Bartholomew's GSMA Mobile Coverage Explorer database is the source for a derived raster, 1km resolution mobile broadband coverage - 1/0 (coverage/no coverage) - dataset. The grid results from the selection and aggregation of 3G (strong signal), 4G and 5G grids, from both operator submissions and OpenCell ID[8] cell tower locations datasets, by applying the following processing steps:

1) Data extraction strong signal coverage 3G strong signal (good indoor coverage) and 4G - MCE_GH3G_2020.tif.
2) Sum (OCI 3G + MCE 3G) and 4G (*r.series* tool cell statistics).
3) Convert raster to Boolean – coverage/no_coverage - (raster calculator).
4) *Clip Raster by mask layer* – clips data by countries borders.
5) Convert raster to vector (*polygonize* tool) - deleted 0 value polygons (no coverage).
6) Check polygons topology (*fix geometries* tool).
7) Create coverage polygon (*dissolve* tool).
8) *Clip Raster by Mask Layer* – top score grids are extracted using the coverage polygon as exclusive criteria.

5.5.3 Urban Areas

Demand/market proximity is a major locational factor for production intensification, where the commercial/specialized urban and peri-urban feedlot production systems can develop.

The urban areas layer is produced using the population density worldpop2020 dataset, the dataset is obtained following several steps:

1) *Raster calculator* - selection of all cells above the 1500 habitants per km² (1/0) layer – Pop2020>1500.
2) *Polygonise too*l - Converts contiguous pixel cells into polygons.
3) *Select by geometry* - all polygons with an area larger than 25 km²
4) *Buffering* – Computes a buffer area for all the features in an input layer using fixed or dynamic distance – 10km.

[8] OpenCellID is a collaborative community project that collects GPS position of user equipment to locate cell towers. As of October 2017, the database contained almost 36 million unique GSM Cell IDs. With more than 75.000 contributors and millions of daily measurements.

5) *Clip Raster by Mask Layer* – top score grids are extracted using the urban areas polygon as exclusive criteria.

5.5.4 Final Location Mapping

After cliping selected top score locations using exclusive criteria, a visual map inspection was performed to detect: logistics and industrial districts, port and railway station proximity, availability of space when in urban area and avoiding residential and touristic areas or districts.
Final location maps put in evidence the livestock production systems and the weight of the dairy herd density, recommended sites are pinpointed based on a detailed large scale analysis of OpenstreetMap, Google and Bing maps and satellite imagery.
Although it involves value judgement, this exercise step tests a return from the model abstraction to field reality, can be considered a first step into project operationalization and also allows reducing the number of proposed sites by aggregating close locations or areas. Large scale maps can be produced detailing each recommended location.

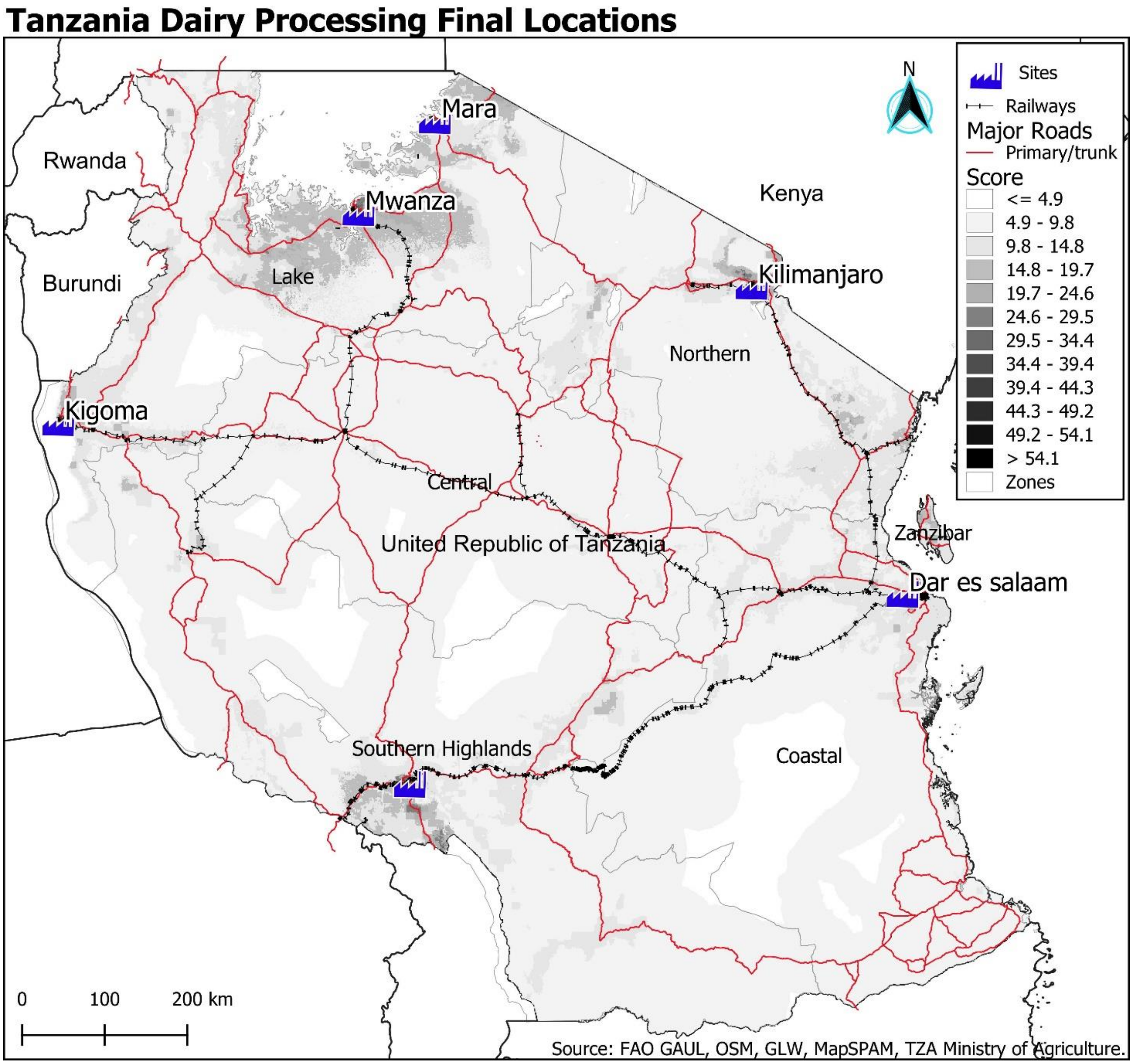

Figure 5 - Dairy recommended processing sites

Region	District	Zones	Location
Dar es salaam	Ubungo MC	Coastal	T1, Morogoro Main Rd, Dar es Salaam
Mwanza	Ilemela MC	Lake	T4, Mwanza
Kigoma	Kigoma Municipal-Ujiji	Lake	Kigoma

Mara	Musoma MC	Lake	T17, Musoma
Kilimanjaro	Moshi MC	Northern	Moshi
Mbeya	Mbeya MC	Southern Highlands	T1, Tazam Rd, Uyole, Mbeya

Table 2- Dairy processing recommended sites

Half of the suggested sites are in the Lake zone with Dar es Salam in the Coastal and two locations in the highlands.

After generating the final location map the georeferenced processing facilities layer was added to check the analysis output against real existing locations. Although the result is not surprising some cases are worth to be further analyzed.

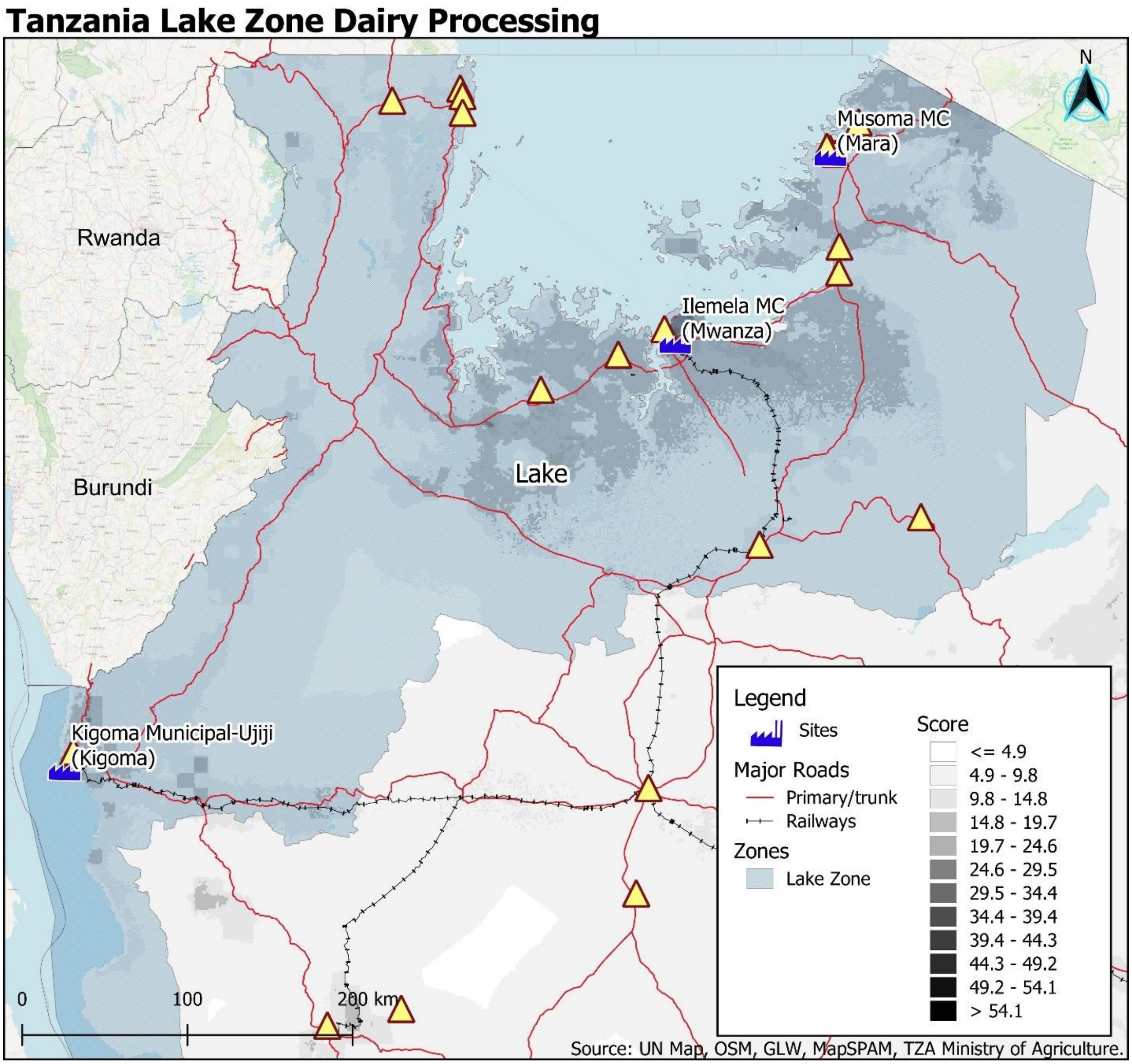

Figure 6 - Lake zone processing sites

Lake production zone tops the number of suggested sites and has several existing processing units in Mara, Mwanza and Kigoma regions but dairy industry data and background check reveals a struggling dairy sector, specifically in Mara region.

Kigoma municipal – Ujiji can be a significant location due to the Tanganyika Lake port and potential Congo DR export market, even more relevant as the model does not weight cross border trade, accessibility to large regional cities. In this site one single processing industry can be found Viemka Agri Enterprise with a low daily processing capacity of 500 which was being underutilized by 60% in 2018.

Industries in Mwanza region were also underperforming, Mother Dairy Sengerema being the largest with 1600 l/day processing capacity but using just about 18.75%.

In Mara region, Musoma, medium and large capacity processors could be found operating at very low levels or closed, the case of Mara Milk (16,000 l/day) and Musoma Dairy (120,000 l/day). The region has one of the largest installed capacity but also with its lowest utilization (Lunogelo et al., 2020), it was considered “the hub of dairy industry in early 2000s” involving thousands of smallholder farmers and small and medium scale dairy processing facilities, some of which collapsed. According to the *Research on SME Inclusion in the Dairy Industry in Tanzania* (Lasway, 2020) possible reasons could be related to : (i) limited artificial insemination services (Ais) (ii) mismanagement of funds and (iii) limited milk collection system and cooling centers, most currently non-operational.

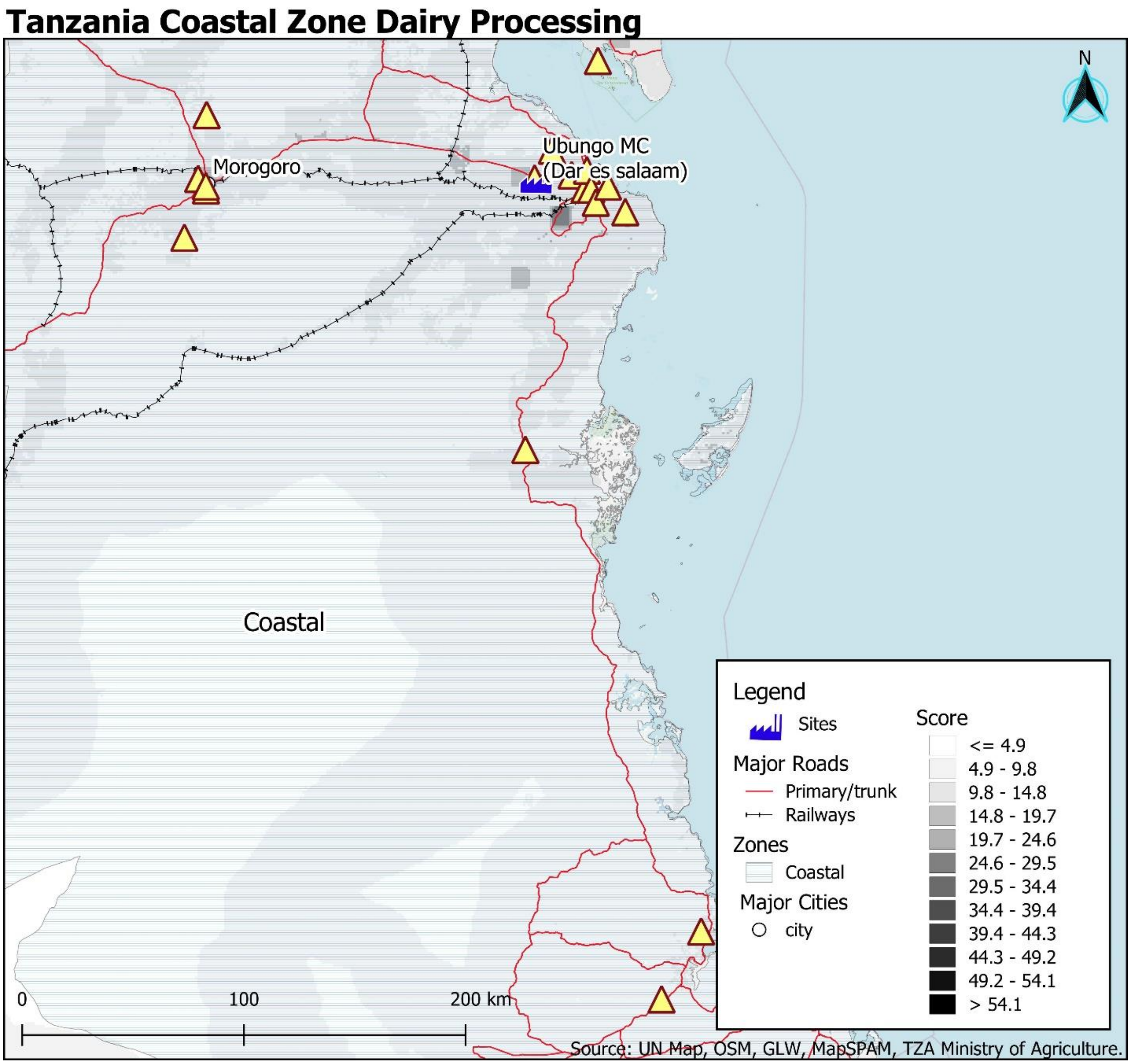

Figure 7 - Coastal zone processing sites

Predictably the northern coastal zone regions have higher location scores than the south and Dar es Salam figures as the single recommend site even though the population size was not weighted. The fact that it is possible to find a processing facility - Ubungu Chawakimu Cooperative – at a very short distance from the proposed site also concurs to sanction the applied methodology, but the country's economic capital and most populated city has an obviously intensification potential, especially if taking into account accessibility to output markets.

A cluster of processing industries can be found in Morogoro in the north Morogoro Region, although the area does not score in the top for processing sites.

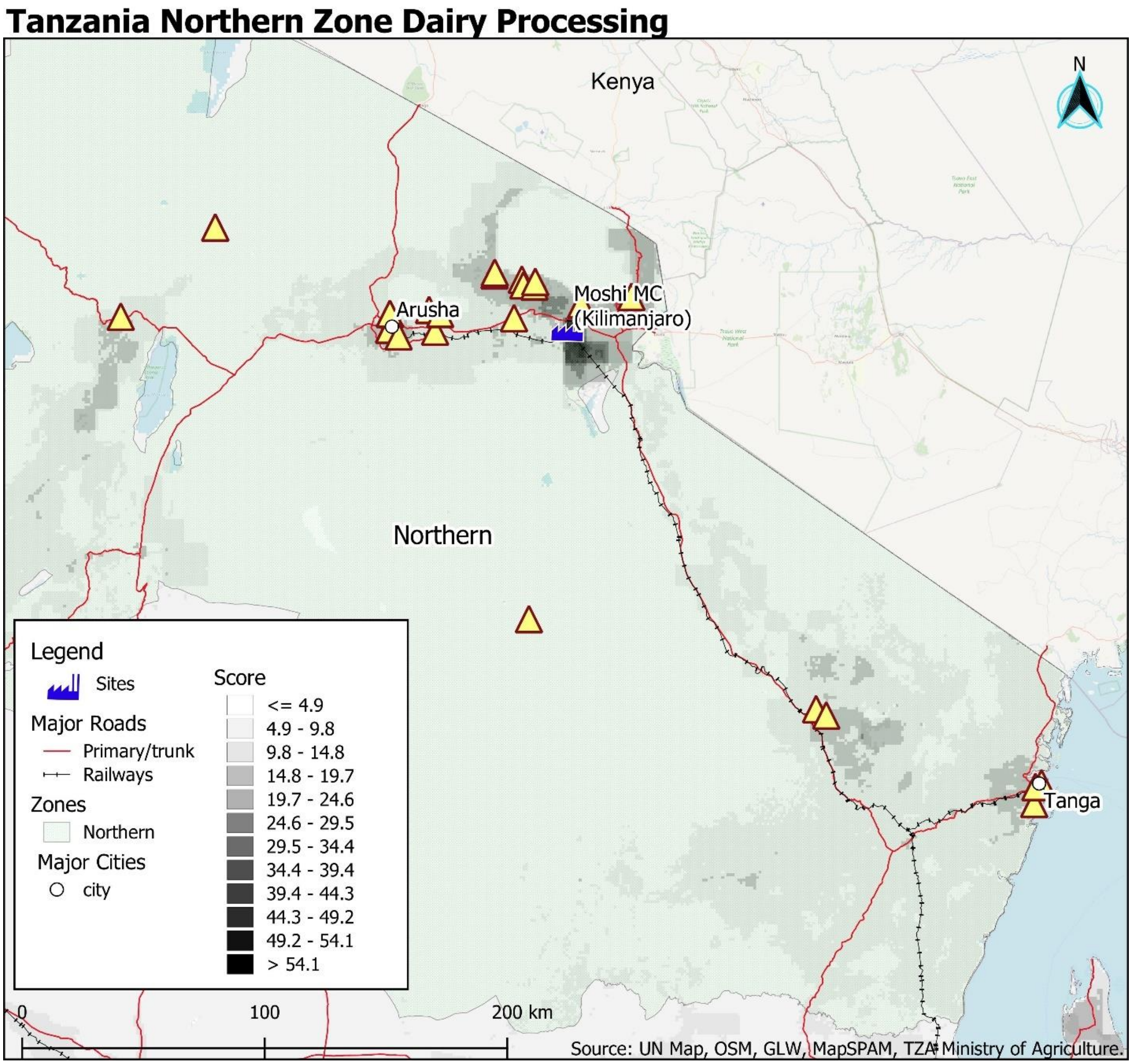

Figure 8 - Northern zone processing sites

The northern zone includes the highland regions Arusha, Kilimanjaro and Manyara (north) and Tanga in the coastal production zone. In most of the northern zone smallholder dairying with grade cattle dominates and can be found in in urban, peri-urban and rural areas (Nell et al., 2014). For natural reasons Kilimanjaro and Arusha concentrate a large portion of the crossbred cattle

population and there is clear evidence of that in the high number of processing facilities, a total processing installed capacity of around 123,900 l/day, more than 40% of the national.
Despite 28 processing industries in northern highlands, none is based in Moshi where the analysis points the highest potential site.

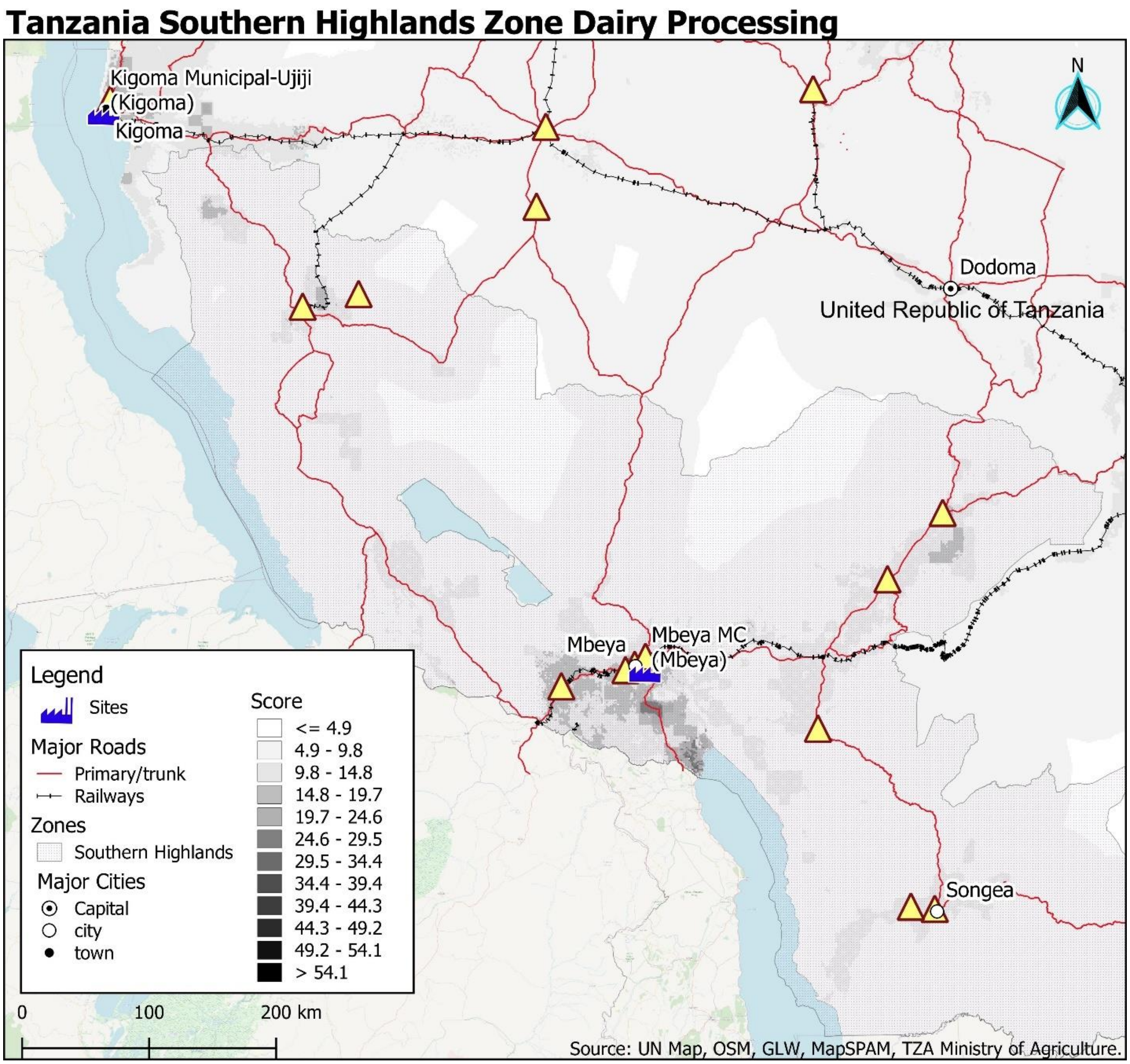

Figure 9 - Southern highlands zone processing sites

Finally, in the southern highlands Mbeya region, Mbeya MC Uyole, is a top score location satisfying all the exclusive criteria. It can be considered a strategic location in what regards surplus export potential since it is the closer to Malawi, Mozambique and Congo DC and is served by the Itungi Port (Lake Malawi), again, this fact would be reinforced if the model weighted cross border trade potential/accessibility to regional cities.

Conclusions

Agriculture is the main economic sector in Tanzania and livestock keeping is also practiced in urban and peri-urban areas. Milk is produced by poor, small-scale livestock farmers that trade primarily through informal channels. Per capita consumption is low and with a close to irrelevant share of it being processed. Adding to that setting there is low financial and credit services coverage, poor milk collection and distribution, and high incidence of livestock diseases.

Demand potential for dairy processed products is high but, *business as usual* scenario, projections point to a large supply deficit.

Based on an understating of favorable natural conditions, national policies point to investments in genetics, health, and nutrition, aiming to close consumption requirement and to generate surplus, thru dairy cow productivity growth, the expansion of the dairy herd, enhancement of the processing industry and marketing. The stated goal is to convert traditional family dairy farms into market-oriented family dairy systems. Among the policy priorities is the investment in commercial milk production and processing.

Tanzania livestock production systems are classified into three livestock production zones, Central, Coastal &Lake, and Highlands. In addition, there is a commercial/specialized livestock, the urban and peri-urban feedlot, production system, across all zones.

Geography imposes genetic, feed, and health conditions that lead to the assumption that intensification is more likely to occur in the Lake&Coastal and Highlands production zones.

Our initial research question: *What would be the best site to locate a dairy processing unit (UHT and milk powder) in Tanzania?* Is answered with background and context sector literature review and by applying a GIS-MCDA methodology to model supply, demand and infrastructure/accessibility.

The theory is:

First, large dairy processing (UHT and milk powder) industries need a bulky volume of raw milk. Consequently, input/supply location criterion is defined by production intensification potential regions. Production (crop/livestock) spatial data is used to identify geographical areas where: additional feed exists or can be produced, crossbred dairy cattle finds conditions to prosper and, a high dairy herd density is already present.

Second, large dairy processing facilities will tendentially locate close to output markets. Large cities, urban areas, and high human population density are therefore our demand/output criteria.

Third and final, UHT and/or milk powder plant location is also dependent on good logistic conditions, accessibility to output market, that criterion is translated to the accessibility to large cities.

Spatial decision involves a set of alternatives and multiple assessment criteria. GIS-MCDA proposes a method to convert and combine spatial data/geographical information and decision-makers criteria to attain evidence, it provides a replicable model, improves communication, offers diverse problem and solution standpoints, and helps refining specification and/or criteria. From a critical standpoint it can be stated that, while data analysis and evidence gathering through GIS modeling can contribute to support decision-making processes, in reality, a complex set of socio-economic, political, cultural, ethno-anthropological aspects and power relations shape and govern most decision-making processes.

Modeling is also as good as the input data, its quality and reliability support the extent to which conclusions can be trusted, and these are just as sound as the analysis conducted. From that prism, specification and objectives define the modeling assumptions and approximations and can always produce distinct answers.

Dairy processing location score as a measure of intensification potential clearly identifies Lake and Highlands larger high score areas and production zones, while Coastal Tanga, Dar es Salam, Zanzibar would benefit from market (demand) proximity and high accessibility.

Lake production zone tops the number of suggested sites. There are several processing units in Mara, Mwanza and Kigoma regions but dairy industry data and background check reveals a struggling dairy sector. Predictably, the northern coastal zone regions have higher location scores than the south and Dar es Salam figures as the single recommend site even though the population size was not weighted on the city accessibility layer. In the northern zone highland regions of Arusha, Kilimanjaro and Manyara (north) and Tanga in the coastal zone, smallholder dairying with grade cattle dominates. Kilimanjaro and Arusha concentrate a large portion of the crossbred cattle population and a high number of processing facilities and installed capacity, Moshi is the highest potential site. Finally, in the southern highlands Mbeya region, Mbeya MC Uyole, is a top score location satisfying all the exclusive criteria. It can be considered a strategic location in what regards surplus export potential since it is the closer to Malawi, Mozambique and Congo DC and is served by the Itungi Port (Lake Malawi), again, this could be reinforced if the model weighted cross border trade potential/accessibility to regional cities.

Finally, caution and examination should be considered over possible impacts of dairy production intensification and industrialization. Sector industrialization leads to vertical and horizontal concentration, attracts large multinational firms, and changes production spatial patterns. This transformation, while increasing production, increasing tax returns by potentially formalizing trade, can also have adverse nonplanned consequences on livelihoods, particularly for a small-scale/smallholder family-based production fabric, but also can produce negative environmental impacts from pressure on input resources and output pollution effluents and GHG.

Closing Remarks

This closing remarks discuss limitations of the findings, changes in methodology and additional location criteria, closing with a few points on future directions.

As with any modelling exercise, assumptions in criteria selection and weighting determine the output. GIS data capture, analysis, processing and representation involve a high degree of generalization and derived, calculated and classified values, with consequent loss of information and detail. There are obvious pitfalls in modeling abstraction reductionism that must be taken into consideration when in a decision-making or operationalization phase.

Assumptions on the supply side should be noted:

A. A dairy intensification input grid weights dairy herd density, crop production and livestock production systems, this can somehow be redundant, selected livestock production systems already define areas where the length of growing period consents crops and forage production allowing a higher density of dairy herd.
B. A high dairy herd density can also indicate an already existing dairy processing operation.

Some of the assumptions in travel time/cost (accessibility) modeling must be emphasized:

A. The defined speed/cost input parameters for each transportation mode.
B. Road speed is considered for cargo and as single value for the considered road classes: 'motorway', 'truck', 'primary' and 'secondary'[9].
C. River navigation is not modeled for Tanzania.
D. Accessibility to ports is not modeled. Port lakes can have some relevance if considered accessibility to regional cities, since identified export markets, neighboring countries, Congo D.R., Malawi, Mozambique, can be reached through Lake Victoria, Lake Tanganyika or Lake Malawi.

[9] See (Ramm, 2019) for details on road classification.

E. Accessibility to cities - Cross border trade (demand), accessibility to large regional cities is not considered since Kenya, Burundi Uganda and Rwanda have strong dairy sectors, and Congo DR, Malawi, and Mozambique can become import partners in the medium to long term. Nevertheless, the sector still has a production consumption deficit.

Some assumptions and pitfalls on demand factors must also be marked.

A. Population density – A high concentration of population in metropolitan areas implies a frequency distribution extremely left skewed. A classification method rather than normalization can allow an informed decision on the definition on urban/rural settlement typology thresholds.

Distinct criteria aggregation requires data classification or normalization/scaling with consequent loss of data.

The final location maps geoprocessing also involves a degree of generalization. It is based on the definition of a threshold linear distance to/from banks and roads, with several physical and human geographical singularities potentially impacting on results accuracy and reliability. There are also assumptions involved in the estimation of mobile broadband coverage maps.

In this version of the GIS-MCDA final location/sites are pinpointed using visual inspection of large-scale available open data and mapping applications. This value judgement assessment uses as criteria: (i) port and railway station proximity; (ii) logistics and industrial districts/areas; (iii) availability of space. Although it involves the analyst value judgement, the exercise tests a return from the model abstraction to field reality, can be considered a first step into operationalization and allows reducing the number of proposed sites by aggregating close locations or areas.

Part of the geoprocessing was semi-automated through the development of algorithms (python scrips), namely:

- Preprocessing, extraction and editing of vector layers (*AuxData* script).

- Computation of accessibility maps cumulative travel time/cost from ports (*TravelCostTimeSurface script*).

Scripts can also be developed for other geoprocessing steps:

- Final top score map - Sequence of geoprocessing functions for IT access, road buffer distance, mobile broadband connectivity, and urban area buffer.

Full automation of the geoprocessing is not realistic as distinct geographies impose demand, supply and infrastructure specific characteristics: population distribution and settlement typology, agroecological zones, livestock and farming production systems, crop commodities and seasonality, infrastructural quality and development, navigation of inland waters and rivers.

Some possible developments include methodological enhancement of current features or the introduction of new criteria:

1. Infrastructure and accessibility dimension:
    a. Railways are modeled as linear phenomena; railway connectivity nevertheless imposes a tunnel effect with time-distance compression between points (stations) but low accessibly in between; Motorways can also be modeled using access points.
    b. Road network modeling uses a single value for cargo irrespective of road class (motorway, truck, primary, secondary) or road speed limits.
    c. Accessibility to regional cities can be included to include cross border, potential surplus trading to Congo DR, Malawi and Mozambique.
2. Socioeconomic data – Other subnational scale level data like poverty incidence.
3. Production dimension:
    a. Other livestock sector value chains: poultry; red meat; eggs; leather.
4. Milk collection and chilling centers – Milk processors depend on stable and reliable milk supply. Limited milk collection system and cooling centers is mentioned as one of the causes for Mara region dairy sector stagnation and decadence. MCCs also contribute to strengthen smallholder associations and bargaining power and by doing so can be

expected to bring development benefits to rural poor livestock dependents. Because MCCs must be in much higher numbers and scattered over larger areas, modeling specification need to address distinct business rules, but basic input data is already in place.

The model has flexibility to accommodate new developments or improvements or can be closed with the present configuration to be automated and to process a larger number of countries and regions. Different scenarios and outputs can be explored using the location criteria weighting (weighted sum) as a *what if scenario* tool.

Bibliography

AscoughII, J. C., Rector, H. D., Hoag, D. L., McMaster, G. S., Vandenberg, B. C., Shaffer, M. J., Weltz, M. A., & Ahjua, L. R. (2019). Multicriteria Spatial Decision Support Systems: Overview, Applications, and Future Research Directions. *Extended Education – an International Perspective*, 175–180. https://doi.org/10.2307/j.ctvdf0hzj.10

Boroushaki, S., & Malczewski, J. (2010). Measuring consensus for collaborative decision-making: A GIS-based approach. *Computers, Environment and Urban Systems*, *34*(4), 322–332. https://doi.org/10.1016/j.compenvurbsys.2010.02.006

Dalberg, Tanzania Dairy Board, Ministry of Liverstock and Fisheries, & Tanzania National Bureau of Statistics. (2019). *Sector Analysis: Price, Cost, Revenue, Profitability and Human capital chains in Tanzania's Dairy Sector*.

Food and Agriculture Organization of the United Nations (FAO), & New Zeland Agricultural Greenhouse Gas Research Centre. (2019). *OPTIONS FOR LOW-EMISSION DEVELOPMENT IN THE TANZANIA DAIRY SECTOR Reducing enteric methane for food security and livelihoods*. http://www.fao.org/3/CA3215EN/ca3215en.pdf

Gilbert, M., Nicolas, G., Cinardi, G., Van Boeckel, T. P., Vanwambeke, S. O., Wint, G. R. W., & Robinson, T. P. (2018). Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Scientific Data*, *5*, 1–11. https://doi.org/10.1038/sdata.2018.227

ILRI, & CGIAR. (2017). Tanzania Livestock Master Plan (2017/2018 - 2021/2022). In *United Republic of Tanzania, Ministry of Livestock and Fisheries*. https://www.mifugouvuvi.go.tz/uploads/projects/1553601793-TANZANIA LIVESTOCK MASTER PLAN.pdf

ILRI, CGIAR, & MLF-URT. (2017). *Tanzania Livestock Sector Analysis (2016/2017- 2032/2033)*. agrichub.org/media/127929_MOFL_Tanzania_LSA_compressed.pdf

Kitchin, R. (2014a). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 205395171452848. https://doi.org/10.1177/2053951714528481

Kitchin, R. (2014b). *The Data Revolution: BIG data, open data, data infrastructures & their consequences*. SAGE.

Kurwijila, R. L., Omore, A., & Grace, D. (2012). The Tanzania Dairy Industry. *Sokoine University of Agriculture*, 1–6. https://cgspace.cgiar.org/bitstream/handle/10568/33865/TanzaniaDairyIndustryOverview2012.pdf?sequence=1

Lasway, J. (2020). *On-going Research on SME Inclusion in the Dairy Industry in Tanzania*.

Innovation and Inclusion in Agro-Processing. https://iiap.info/on-going-research-on-sme-inclusion-in-the-dairy-industry-in-tanzania/

Lunogelo, H., Makene, F., & Gray, H. (2020). Dairy Processing In Tanzania Prospects for SME inclusion. In *Innovation & Inclusion in Agro-processing* (Issue April). https://doi.org/10.3362/9781780444000

Malczewski, J. (2006). GIS-based multicriteria decision analysis: A survey of the literature. *International Journal of Geographical Information Science*, *20*(7), 703–726. https://doi.org/10.1080/13658810600661508

Mulrooney, T., Beratan, K., McGinn, C., & Branch, B. (2017). A comparison of raster-based travel time surfaces against vector-based network calculations as applied in the study of rural food deserts. *Applied Geography*, *78*, 12–21. https://doi.org/10.1016/j.apgeog.2016.10.006

Nell, A. J., Schiere, H., & Bol, S. (2014). QUICK SCAN Dairy Sector TANZANIA. *Dutch Ministry of Economic Affairs (Dept. of European Agricultural Policy and Food Security; DG Agro)*, *January*. https://edepot.wur.nl/334382

NIRAS. (2010). Survey on Dairy Products Market in Tanzania - Final Report. In *Niras* (Issue May). http://www.tzdpg.or.tz/fileadmin/_migrated/content_uploads/Final_Dairy_Market_Report_12_1_.5.2010.pdf

Njombe, A.P., Yakobo Msanga Nathaniel Mbwambo, N. M. (2011). the Tanzania Dairy Industry : Status, Opportunities, and Prospects. *The Tanzania Dairy Industry: Status, Opportunities and Prospects*, *May*, 25–27.

Pica-Ciamarra, U., Baker, D., Chassama, J., Fadiga, M., & Nsiima, L. (2011). *Linking smallholders to livestock markets in Tanzania Combininng Market and Household Survey Data*. World Bank, FAO, AU-IBAR, ILRI. https://openknowledge.worldbank.org/bitstream/handle/10986/17888/866410WP0PAP0a00Box385182B00PUBLIC0.pdf;sequence=1

Quaedackers, P., Linden, V. Van Der, & Boer, D. De. (2009). *Organizing milk collection in the Tanzanian dairy sector* (Issue September). http://www.tzdpg.or.tz/fileadmin/_migrated/content_uploads/TAMPA_report_FINAL_Draft.pdf

Ramm, F. (2019). *OpenStreetMap Data in Layered GIS Format V7* (p. 21). Geofabrick.

Robinson, T., Thornton, P., Franceschini, G., Kruska, R., Chiozza, F., Notenbaert, A., Cecchi, G., Herrero, M., Epprecht, M., Fritz, S., You, L., Conchedda, G., & See, L. (2011). *Global Livestock Production Systems*. Food and Agriculture Organization of the United Nations (FAO), International Livestock Research Institute. http://www.fao.org/3/i2414e/i2414e.pdf

Tanzania National Bureau of Statistics. (2017). Annual Agriculture Sample Survey Crop and

Livestock Report. *Annual Agriculture Sample Survey Crop and Livestock Report*, 1–181. https://www.nbs.go.tz/nbs/takwimu/Agriculture/2016-17_AASS Report _Final.pdf

Wood-Sichra, U., Joglekar, A. B., & You, L. (2016). Spatial Production Allocation Model (SPAM) 2005: Technical Documentation. In *Harvest Choice*. http://mapspam.info

Annex – TravelCostTimeSurface algorithm model diagram