# Forecasting The UN Sustainable Development Goals

Yassir Alharbi[1,3][0000-0001-6764-030X], Daniel Arribas-Bel[2][0000-0002-6274-1619], and Frans Coenen[1][0000-0003-1026-6649]

[1] Department of Computer Science, The University of Liverpool, Liverpool L69 3BX, United Kingdom
[2] Department of Geography and Planning, The University of Liverpool, Liverpool L69 3BX, United Kingdom
[3] Almahd College, Taibah University Al-Madinah Al-Munawarah, Saudi Arabia
yassir.alharbi, coenen, d.arribas-bel@liverpool.ac.uk

**Abstract.** This paper presents a review and in-depth analysis of the Sustainable Development Goal Track, Trace, and Forecast (SDG-TTF) framework for UN Sustainable Development Goal (SDG) attainment forecasting. Unlike earlier SDG attainment forecasting frameworks, the SDG-TTF framework considers the possibility for causal relationships between SDG indicators, both within a given geographic entity (intra-entity relationships) and between the current entity and its neighbouring geographic entities (inter-entity relationships). The difficulty lies in identifying such causal linkages. Six different mechanisms are considered. The discovered causal relationships are then used to generate multivariate time series prediction models within a bottom-up SDG prediction taxonomy. The overall framework was assessed using three different geographical regions. The results demonstrated that the Extended SDG-TTF framework was capable of producing better predictions than competing models that do not account for the possibility of intra and inter-causal linkages.

**Keywords:** Time series causality · Missing values· Hierarchical classification · Time series forecasting · Sustainable Development Goals.

## 1 Introduction

On 8 September 2000, at the end of the United Nations (UN) Millennium Summit, world leaders adopted eight Millennium Development Goals (MDGs) to be achieved before 2015. The goals are listed in Table 1. Most of the specified goals were achieved by most countries [28], and the MDG initiative was declared to be a success.

The success of MDG initiative paved the way for another set of goals. In September 2015, the UN introduced the Sustainable Development Goals (SDGs), listed in Table 2, to be achieved by 2030 [27]. However, this time the goals covered a broader range of domains. The vision was that achieving these goals would

| |
|---|
| 1. To eradicate extreme poverty and hunger. |
| 2. To achieve universal primary education. |
| 3. To promote gender equality and empower women; |
| 4. To reduce child mortality. |
| 5. To improve maternal health. |
| 6. To combat HIV/AIDS, malaria, and other diseases. |
| 7. To ensure environmental sustainability. |
| 8. To develop a global partnership for development. |

**Table 1.** The eight 2000 Millennium Development Goals (MDGs) [3]

| |
|---|
| 1. No Poverty. |
| 2. Zero Hunger. |
| 3. Good Health and Well-being. |
| 4. Quality Education. |
| 5. Gender Equality. |
| 6. Clean Water and Sanitation. |
| 7. Affordable and Clean Energy. |
| 8. Decent Work and Economic Growth. |
| 9. Industry, Innovation and Infrastructure. |
| 10. Reduced Inequality. |
| 11. Sustainable Cities and Communities. |
| 12. Responsible Consumption and Production. |
| 13. Climate Action. |
| 14. Life Below Water. |
| 15. Life on Land. |
| 16. Peace and Justice Strong Institutions. |
| 17. Partnerships to Achieve the Goals. |

**Table 2.** The seventeen 2005 Sustainable Development Goals (SDGs)

provide for a world free from hunger and poverty and ensure the sustainability of natural resources and the protection of the environment. The philosophical underpinning for the SDGs initiative, and the MDG initiative, was the idea that the world is a connected place and that all UN members should therefore work together to ensure the attainment of these goals for all member states [8,22].

Given the foregoing, predicting whether geographic regions will meet their SDGs or not is of significant interest. Of note is work directed at using machine learning to predict SDG attainment [2,4,21,24]. Most of this existing work assumes that the various SDG targets and indicators are independent of one another. However, this is clearly not the case. The SDGs can be categorised into 4 different levels, as shown in Figure 1:

**Biosphere:** environmental-related goals.
**Society:** Goals related to empowering society, such as by: (i) eradicating poverty, (ii) promoting health and equality and (iii) promoting sustainable cities.

**Economy:** Goals related to economic growth using responsible consumption while reducing workforce inequality.

**Partnership:** Goal 17, which exists to promote a global effort to ensure that the SDGs are attained with respect to all countries.

By considering the SDGs in terms of the above categorisation, it can be seen that SDGs are related and connected. For example, without sustained clean water and sanitation (SDG 6), food preparation will be harder; thus, it will affect the attainability of SDG 2 (Zero Hunger); and therefore, SDG 8 (Decent work and economic growth) will not be fulfilled. Given this interconnected nature of the SDGs one can see that there exists a causal relation between individual SDGs.


**Fig. 1.** Interconnectedness in SDG courtesy of Azote and the Stockholm Resilience Centre, Stockholm University [22]

In [3] the SDG Multivariate Track, Trace and Forecast (SDG-TTF) framework was presented that took into consideration both intra-entity relationships and inter-geographic region causalities between SDGs. The proposed SDG-TTF model incorporates the hierarchical framework from [2], and the ACA causality relationship mechanism from [4] for intra- and inter-entity relationship discovery.

This paper presents a much more comprehensive evaluation of the SDG-TTF framework presented in [3] by considering: (i) three different mechanisms for addressing the missing data problem, (ii) three different mechanisms to address the scaling issue that exists within the SDG data, (iii) six different mechanisms for discovering causal relationships and (iv) using a much more substantial portion of the available SDG data to evaluate the framework than originally used in [3], data covering North Africa, East Asia and Northern Europe. The paper also presents a more detailed analysis of the required pre-processing and the adopted mathematical representation.

The rest of this paper is organised as follows. In the following section, Section 2, a brief literature review of relevant work underpinning the work presented in this paper is given. The SDG application domain and the SDG time series data set is described in Section 3 together with the required pre-processing of the SDG data. The SDG-TTF framework is described in Section 4 and its evaluation in Section 5. The framework's operation is given in Section 6. The paper concludes with a summary of the main findings, a number of proposed directions for future research, in Section 7. All the data provided in this paper can be found in the project Github repository[1]

## 2 Literature Review

The proposed SDG-TTF approach addresses two fundamental challenges: (i) short time series forecasting and (ii) time series causal inference. Previous work in these two areas is therefore considered in the first two sub-sections in this literature review. The literature review is completed with some discussion of previous work directed at SDG forecasting.

### 2.1 Short time series forecasting

Short time series forecasting is challenging because it is difficult to perform meaningful out of sample evaluation, or cross validation, given a low number of observations [13]. From the literature a range of methods have been proposed to address this issue, see for example [7]. However, the proposed solutions tend to still insist on 50 or more observations. In the case of the SDG data, the sample size is less than 20 points. The FBProphet time series forecasting tool was used in [2] for the purpose of SDG attainment prediction where it was demonstrated that FBProphet produced a better prediction accuracy over two alternatives, ARMA and ARIMA.

However, FBProhpet is a uni-variate predictor; given that the focus of this paper is prediction using sets of causal-related time series a multi-variate approach is required. A multivariate time series forecasting model, using Long Short Term Memory (LSTM) networks, was presented in [14]. The LSTM model demonstrated a better overall performance compared to ARMA and ARIMA [7]. The LSTM model was adopted in [4] for multi-variate SDG attainment forecasting. More generally, LSTM models have been widely adopted with respect to many real-life applications such as weather [20] and stock market [6] prediction. With respect to the work presented in this paper an Encoder-Decoder LSTM, was used [14]. LSTM typically performs better when large data sets are used.

[1] https://github.com/Yassir-Alharbi/Sustainable-Development-goals

But also seems to perform well when a large number of short time series are available, as in the case of the SDG prediction application considered here.

### 2.2 Time series causal inference

Causal inference is concerned with the process of establishing a connection (or the lack of a connection) between events or instances. Given two candidate time series, $A = \{a_1, a_s, \ldots, a_n\}$ and $B = \{b_1, b_2, \ldots, b_m\}$, where we wish to establish that $B$ is causality-related to $A$, this is typically established using a prediction mechanism that uses the "lag" $\{b_1, \ldots, b_{m-1}\}$ to predict $a_n$. We then compare the predicted value for $a_n$ with the known value, for example using the Root Mean Square Error (RMSE) as a comparison metric. If the two values are close then we can say that the"time series $A$ is causality-related to time series $B$".

There are a number of mechanisms that can be adopted to achieve the above. With respect to the work presented in this paper, six such mechanisms were considered: (i) Granger Causality (GC), (ii) the Temporal Causal Discovery Framework (TCDF), (iii) Pearson coefficient, (iv) Lasso, (v) the Mann-Whitney U Test. and (vi) ACA. Each is discussed in some further detail below.

**Granger Causality** Granger Causality (GC) is one of the most widely used causal inference mechanisms found in the literature [8,17]. It was introduced in the 1960s and is calculated as shown in Equation 1 where: (i) $X$ and $Y$ are time series, (ii) $a$ and $b$ are the laggs of $X$ and $Y$, (iii) $t$ is the current time step and (iv) $e$ is a residual error. The idea is that if time series $X$ "granger causes" time series $Y$, then the past values of $X$ should contain helpful information to forecast $Y$ in a manner that would be better than when forecasting $y$ using only historical data associated with $Y$. The variation of GC that was used with respect to the research presented in this paper is the Stats-models variation [23]. GC has been used previously in the context of SDG prediction, for example in [8] 20,000 pairs of time series that featured causal relationship were found.

$$Xt = a_1 X_{t-1} + b_1 Y_{t-1} + e \tag{1}$$

**Temporal Causal Discovery Framework** The Temporal Causal Discovery Framework (TCDF) [18] is an alternative mechanism to GC to determine whether a time series $A$ has a causal association with a time series $B$. TCDF uses a Convolutional Neural Network (CNN) whose internal parameters are interpreted to discover causal relations. The framework has been shown to not work well with respect to short time series. For best performance it is suggested that 1000 data points are required, but is still considered in this paper.

**Pearson Correlation** Pearson Correlation [10] has been used to measure the correlations between any given pair of time series. The mechanism assumes linearity of the data. This assumptions holds with respect to many SDG time series that are typically linearly spaced, and therefore seems an appropriate choice.

**Lasso** Lasso [26] is an L1 regularisation technique frequently used to reduce high dimensionality data, which can also be employed to establish the existence of a causality between variables [9,26]. LASSO reduces the dimensionality of the input data set by penalising variances to zero, thus allowing irrelevant variables to be removed. Equation 2 shows the LASSO cost function. Inspection of the equation indicates that the first part is the *squared error* function, whilst the second part is a penalty applied to the regression slope. If $\lambda$ is equal to 0, then the function becomes a normal regression. However, if $\lambda$ is not 0 coefficients are penalised accordingly, leaving only coefficients that can explain the variance in the data.

$$L\ C\ F = \sum_{i=1}^{n} \left( y_i - \sum_{j} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{2}$$

**Mann-Whitney U Test** The Mann-Whitney U Test [1] is the fifth causal inference mechanism used in this paper. The test is used to determine if any two pairs of time series are statistically different. It is a non-parametric test (unlike, for example, Lasso).

**ACA** The last of the six causality discovery mechanisms considered in this paper is the ACA mechanism proposed in [4]; the name is derived from the author's initials. Essentially this is an ensemble of the above five mechanisms which was found to outperform the above mechanisms when used individually.

### 2.3 Sustainable Development Goals Forecasting

Previous work directed at the forecasting of SDG attainment can be divided into two main categories: (i) single target forecasting or (ii) multiple target forecasting. The first is directed at forecasting with respect to an individual SDG or specific geographical location. Much existing work falls into this category. Examples can be found in [21] and [24] where forecasting was directed at a a specific SDG (electricity supply) and specific region (Ukraine) respectively. The second is concerned with predicting multiple targets. Examples of this second approach include the SDG-AP and SDG-CAP frameworks presented in [2] and [4] respectively that were referenced in the introduction to this paper and that are used for comparison purposes with respect to the evaluation of the SDG TTF framework given later in this paper.

## 3 The SDG Data Set and Associated Data Preparation

To maintain oversight of the SDG agenda, the UN periodically releases SDG related data on the www platform of the United Nations Department of Economic and Social Affairs' Statistics Division[2] . Once on the SDG data website, the data can be downloaded, partially or wholly, in a CSV format. The SDG platform holds data related to 346 countries. In addition the platform features collated data for regional groupings, such as Sub-Saharan Africa, Northern Africa, Western Asia, Central and Southern Asia, and so on. The total number of indicators is 561 divided across 169 targets (and 17 SDGs). Each indicator will have one or more sub-indicators.

The SDG data set comprises a set of records $\{R_1, R_2, \ldots\}$. Each record $R_i$ comprises a set of values $\{v_1, v_2, \ldots\}$ where each value corresponds to a set of attributes $A = \{a_1, a_2, \ldots\}$. The attributes take either a categorical or numerical value. Thus $D$ comprises a single, very large, table with the columns representing a range of numerical and categorical attributes and the rows representing single observations related to individual SDG indicators. Each record $R_i$ is date stamped. The set $A$ (the columns in the table) represent the complete set of attributes for all 561 indicators. However, for any one indicator only a small sub-set of the available set of attributes will be relevant. Table 3 gives an example of a SDG record for the country Afghanistan for the year 2015. The example refers to Goal 16 "*Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels*", Target 16.1 is then "*Significantly reduce all forms of violence and related death rates everywhere*", which has associated with it Indicator 16.1.1, "*Number of victims of intentional homicide per 100,000 population, by sex (victims per 100,000 population)*", which in this case has a value of 0.55597 per 100,000 head of population. For ease of reading the table has been arranged in a multi-column format, the record is actually a single row in the SDG database table. From the example it can be seen that many of the attributes are not relevant (such as "Mountain elevation" or "type Of Speed"). This is why the value for many attributes in Table 3 has been set to "NA" (Not Applicable). This is a feature of all the records in the SDG data set.

Note that the indicator reference, 16.1.1 in above the example, incorporates the goal and target references, hence for further processing we only need the indicator reference. In the remainder of this paper we will refer to this as the Goal-Target-Indicator (GTI). Note that targets are also identified, in the raw SDG data, using lower case letters, for example 9.c.1. The GTI is sufficient to identify individual rows if there is only one sub-indicator. However, in many cases we have more than one sub-indicator. Hence, to differentiate between individual time series, a unique "Individual Series" (IS) identifier, made up of the Series Code (the fourth attribute in Table 3, and further identifying characters, was devised. Note, from the table, that the series code is made up of three *text segments* separated by underscore characters. IS values were constructed by adding a fourth text segment.

To allow SDG prediction the SDG data, as described above, needed to be pre-processed into a structured format. In [2] and [4] prediction was facilitated by a hierarchical taxonomy SDGs⇒Targets⇒Indicators⇒Sub-indicators with pre-

[2] https://unstats.un.org/sdgs/indicators/database

dictors at the sub-indicator leaf nodes whose results were passed up the tree and combined level-by-level till the root of the tree was reached and a final prediction arrived at. The same approach was adopted with respect to the SDG TTF framework presented here. The structured format thus also had to facilitate the population of the taxonomy, once generated, with respect to individual countries (geographic regions). An important element of populating the taxonomy was the collation of the time series values to be used to create the predictors to be held at the taxonomy leaf nodes. The mechanism adopted with respect to the SDG TTF framework described here for creating the predictors is what sets it apart from the SDG-AP and SDG-CAP frameworks described in [2] and [4].

| | | | |
|---|---|---|---|
| Attribute | Goal | TimeCoverage | Cities |
| Value | 16 | NA | NA |
| Attribute | Target | UpperBound | Counterpart |
| Value | 16.1 | NA | NA |
| Attribute | Indicator | LowerBound | Disability status |
| Value | 16.1.1 | NA | NA |
| Attribute | SeriesCode | BasePeriod | Education level |
| Value | VC_IHR_PSRC | NA | NA |
| Attribute | SeriesDescription | Source | Fiscal intervention stage |
| Value | Number of victims of intentional homicide per 100,000 population, by sex | National Statistical Organization | NA |
| Attribute | GeoAreaCode | GeoInfoUrl | Food Waste Sector |
| Value | 4 | NA | NA |
| Attribute | GeoAreaName | FootNote | Freq |
| Value | Afghanistan | NA | NA |
| Attribute | TimePeriod | Activity | Frequency of Chlorophyll-a concentration |
| Value | 2015 | NA | NA |
| Attribute | Value | Age | Grounds of discrimination |
| Value | 0.55597 | NA | NA |
| Attribute | Time_Detail | Cause of death | Hazard type |
| Value | 2015 | NA | NA |
| Attribute | IHR Capacity | Mode of transportation | Observation Status |
| Value | NA | NA | NA |
| Attribute | Level of requirement | Mountain Elevation | Parliamentary committees |
| Value | NA | NA | NA |
| Attribute | Level/Status | Name of international institution | Policy Domains |
| Value | NA | NA | NA |
| Attribute | Location | Name of non-communicable disease | Policy instruments |
| Value | NA | NA | NA |
| Attribute | Nature | Quantile | Migratory status |
| Value | C | NA | NA |
| Attribute | Report Ordinal | Substance use disorders | Type of speed |
| Value | NA | NA | NA |
| Attribute | Reporting Type | Type of occupation | Type of support |
| Value | G | NA | NA |
| Attribute | Sampling Stations | Type of product | Type of waste treatment |
| Value | NA | NA | NA |
| Attribute | Sex | Type of skill | Units |
| Value | FEMALE | NA | PER_100000_POP |

**Table 3.** SDG Example Record

A schematic of the adopted pre-processing mechanism is given in Figure 2. From the figure it can be seen that the mechanism comprised two stages.

**Stage 1** Taxonomy Generation.
**Stage 2** Missing Value Imputation and Scaling and Generation of Country Data Files.

Each stage is discussed in further detail in the following three subsections.

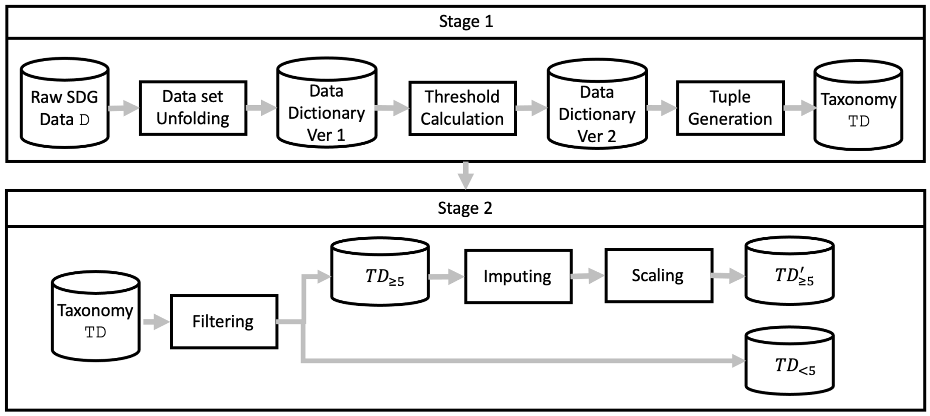


**Fig. 2.** Preprocessing Schematic

### 3.1 Taxonomy Generation (Pre-processing Stage 1)

Stage 1, as noted above, comprises taxonomy generation. The taxonomy, although describing a tree structure, is actually stored as a set of tuples in a data set $TD = \{TR_1, TR_2, \ldots\}$, where each $TR_i \in TD$ is a tuple of the form:

$$TR_i = \langle Geographical\_Region, GTI, IS, T \rangle \tag{3}$$

where: (i) "Geographical_Region" is the name of the country or region of interest, (ii) GTI is the relevant Goal-Target-Indicator, (iii) IS is the relevant Individual Series identifier for the time series held at the leaf node in the topology for a given region, and (iv) $T$ is the time series associated with the leaf node, $T = [t_1, t_2, \ldots]$. Note that the particular relevance of the format of $TD$ is that it allows comparison between records within a given geographic region, and comparison across geographic regions, a central feature of the proposed Extended SDG-TTF framework

Pre-processing Stage 1 is thus about transforming the raw SDG $D$ input into a data set $TD$. The process commences by applying a depth-first "unfolding" operation to the raw data $D$ so as to collect the topology paths for particular region and indicator pairs. The path information is stored in a data dictionary, and intermediate data repository between $D$ and $TD$ designed to facilitate the transformation from $D$ to $TD$. An example of this unfolding, using Afghanistan and GTI 16.1.1, is given in Table 3, is given in Figure 3.

A number of example dictionary entries are given below:

1. {*(Goal : 16), (Target : 16.1), (Indicator : 16.1.1), (Series Code : VC_IHR_PSRC), (Series Description : Number of victims of intentional homicide per 100,000 population , by sex (victims per 100,000 population))) (Geo Area Name : Afghanistan), (Units : PER_100000_POP), (Sex : FEMALE)*}

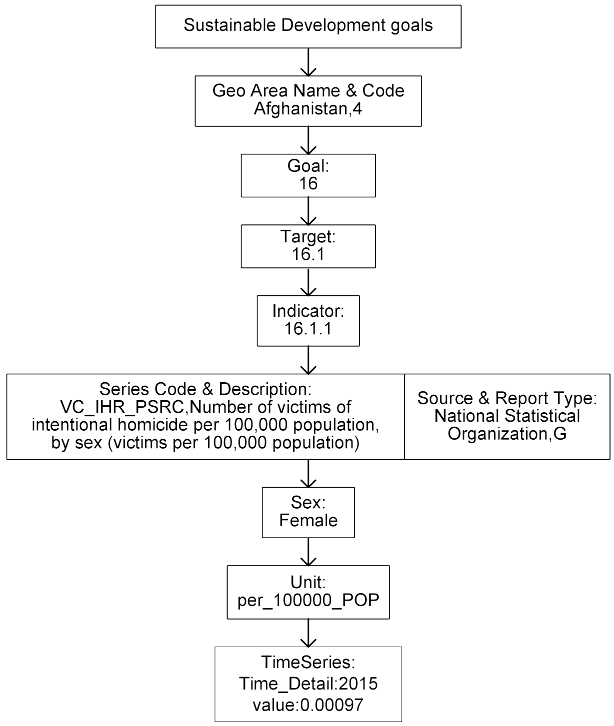


**Fig. 3.** Illustration of the SDG raw data depth-first "unfolding" for the region-GTI pair Afghanistan and 16.1.1

2. {*(Goal : 11), (Target : 13.1), (Indicator : 11.b.2), (Series Code : SG_GOV_LOGV), (Series Description : Number of local governments (number)), (Geo Area Name : Afghanistan), (Units : NUMBER)*}
3. {*(Goal : 15), (Target : 15.4), (Indicator : 15.4.2), (Series Code : ER_MTN_GRNCVI), (Series Description : Mountain Green Cover Index), (Geo Area Name : Afghanistan), (Units : PERCENT), (Observation Status : A), (Mountain Elevation : 4)*}

The first of the above examples is for the region-GTI pair Afghanistan and 16.1.1 used for illustrative purposes in Table 3 and Figure 3. The second two are two additional examples. Note that only the most salient information is stored in the dictionary, the information needed for the topology.

The next stage is to produce a mathematical formulation of the threshold value to be used to determine whether each indicator has been met or not. In some cases this is straight forward, in others this is not so straight forward. For example phrases such as "*Eradicate*" and "*reduce at least by half*" are used to define TGIs. A solution, in the context of the proposed taxonomy, was available in [16] where the authors published guidelines on how to interpret the health target goals from the SDG published Target Goals document, including mathematical definitions. The guidelines in [16] were adopted to translate the textual descriptions of SDG sub-indicators into mathematical ones. We now have all the information needed for our taxonomy.

Table 4 summarises the details of the sub-indicators held in the dictionary with respect to GTIs 1.1.1, 1.2.1 and 1.3.1. In the table, Column 1 gives the GTI, Column 2 gives the "Thresholds" (calculated as described above) that need to be met before the dead line given in Column 3, Column 4 gives the "Series Description", Column 5 the IS indicator and Column 6 the unique ID number for the leaf node. Recall that the IS (Individual Series) indicator is a textual description of sub-indicator and by extension the associate time series. IS identifiers are constructed from the raw data (see Table 3). As noted earlier, the IS indicator comprises four text segments separated by underscore characters. The first three are the series code taken from the raw data (see Table 3). For example, considering the fifth example in Table 4, "VC_IHR_PSRC". The fourth text segment that has been added in this case is "FEMALE" to give an IS indicator of "VC_IHR_PSRC_FEMALE". The unique ID is simply a unique numeric identifier which is simpler to use than the IS indicator (but not at all easy to interpret).

The final step in stage one (see Figure 2) is to transpose the data in the dictionary so that it is held in a set of the form $TD = \{TR_1, TR_2, ...\}$ where each $TR_i \in TD$ is a tuple of the form described above. Note that to reference the time series associated with a record $TR_i$ we will use the notation $T_i$. Table 5 gives a series of example time series for a number of indicators for Afghanistan. Each column represents a time series. The column headings give the relevant GTI and the time series ID. What can be clearly seen from the table is that there are many missing values!

### 3.2 Missing Value Imputation and Scaling (Pre-processing Stage 2)

The theoretical maximum length of any SDG time series is 22 points, covering 22 years of observations from 2000 to 2021. Some indicators, for a small number of countries, have data going back to 1974. Figure 4 shows the number of observations per year with respect to the 30 different countries considered for evaluation purposes in this paper. Inspection of the figure indicates that the majority of the data falls within 2000 and 2018. Beyond 2018 the data is frequently not yet available (in some cases it may never become available). Thus, in the context of the research presented in this paper, only data from 2000 to 2018 was considered; 34,526 time series in total. Therefore, with reference to Figure 2, the first step in Stage 2 was to filter $TD$ so that it comprised only of 18 point time series.

As noted with reference to Table 5, the SDG data set features a significant number of missing values.

The reasons for missing data in the collated SDG time series are varied but can be categorised as either: Missing At Random (MAR) or Not Missing At Random (NMAR) [12]. We can illustrate the distinction by considering the thwo example time series given in Table 6, the first describes the time series

| G.T.I | Thresholds | Date | Series Description | Series Code | ID |
|---|---|---|---|---|---|
| 1.1.1 | <=0.5% | 2030 | Employed population below international poverty line, by sex and age (%) | SI_POV_EMP_15-24_MALE | 1 |
| | | | | SI_POV_EMP_BOTHSEX_15+ | 2 |
| | | | | SI_POV_EMP_BOTHSEX_15-24 | 3 |
| | | | | SI_POV_EMP_BOTHSEX_25+ | 4 |
| | | | | SI_POV_EMP_FEMALE_15+ | 5 |
| | | | | SI_POV_EMP_FEMALE_15-24 | 6 |
| | | | | SI_POV_EMP_FEMALE_25+ | 7 |
| | | | | SI_POV_EMP_MALE_15+ | 8 |
| | | | | SI_POV_EMP_MALE_25+ | 9 |
| | | | Proportion of population below international poverty line (%) | SI_POV_DAY | 10 |
| 1.2.1 | <=50% | | Proportion of population living below the national poverty line (%) | SI_POV_NAHC__ALLAREA | 11 |
| 1.3.1 | >=80% | | [ILO] Proportion of children/households receiving child/family cash benefit (%) | SI_COV_CHLD_BOTHSEX | 12 |
| | | | | SI_COV_CHLD_FEMALE | 13 |
| | | | | SI_COV_CHLD_MALE | 14 |
| | | | [ILO] Proportion of employed population covered in the event of work injury (%) | SI_COV_WKINJRY_BOTHSEX | 15 |
| | | | | SI_COV_WKINJRY_FEMALE | 16 |
| | | | | SI_COV_WKINJRY_MALE | 17 |
| | | | [ILO] Proportion of mothers with new borns receiving maternity cash benefit (%) | SI_COV_MATNL_BOTHSEX | 18 |
| | | | | SI_COV_MATNL_FEMALE | 19 |
| | | | [ILO] Proportion of poor population receiving social assistance cash benefit (%) | SI_COV_POOR_BOTHSEX | 20 |
| | | | [ILO] Proportion of population above statutory pensionable age receiving a pension, by sex (%) | SI_COV_PENSN_BOTHSEX | 21 |
| | | | | SI_COV_PENSN_FEMALE | 22 |
| | | | | SI_COV_PENSN_MALE | 23 |
| | | | [ILO] Proportion of population covered by at least one social protection benefit (%) | SI_COV_BENFTS_BOTHSEX | 24 |
| | | | | SI_COV_BENFTS_FEMALE | 25 |
| | | | | SI_COV_BENFTS_MALE | 26 |
| | | | [ILO] Proportion of population with severe disabilities receiving disability cash benefit (%) | SI_COV_DISAB_BOTHSEX | 27 |
| | | | | SI_COV_DISAB_FEMALE | 28 |
| | | | | SI_COV_DISAB_MALE | 29 |
| | | | [ILO] Proportion of unemployed persons receiving unemployment cash benefit, by sex (%) | SI_COV_UEMP_BOTHSEX | 30 |
| | | | | SI_COV_UEMP_FEMALE | 31 |
| | | | | SI_COV_UEMP_MALE | 32 |
| | | | [ILO] Proportion of vulnerable population receiving social assistance cash benefit (%) | SI_COV_VULN_BOTHSEX | 33 |
| | | | | SI_COV_VULN_FEMALE | 34 |
| | | | | SI_COV_VULN_MALE | 35 |
| | | | [World Bank] Poorest quintile covered by labour market programs (%) | SI_COV_LMKTPQ_ | 36 |
| | | | [World Bank] Poorest quintile covered by social assistance programs (%) | SI_COV_SOCASTPQ_ | 37 |
| | | | | SI_COV_SOCINSPQ_ | 38 |
| | | | [World Bank] Proportion of population covered by labour market programs (%) | SI_COV_LMKT_ | 39 |
| | | | [World Bank] Proportion of population covered by social assistance programs (%) | SI_COV_SOCAST_ | 40 |
| | | | | SI_COV_SOCINS_ | 41 |

**Table 4.** Examples taxonomy leaf node information held in the Data Dictionary

for the indicator "*Direct economic loss attributed to disasters (current United States dollars)*" (GTI 11.5.2, ID 3109). Inspection of the associated time series reveals a time series with only one recorded value, the value of 311. However, the data describes "*loss attributed to disasters*", which by definition (we hope) are not regular occurrences, hence financial losses as a result of disasters are not recorded every year. This type of missing data is thus considered to be NMAR data. The second column describes the time series indicator "*Proportion of population covered by a mobile network, by technology*" (GTI 9.c.1, ID 2712). In this case the absence of the missing data is unclear because Egypt did have mobile services prior to 2014. Thus this type of missing data is considered to be MAR data.

To address the missing data problem the idea was to adopt some kind of data imputation, the process of assigning values to missing attribute value instances according to neighbouring values. This will only work if sufficient neighbouring values are available. Some preliminary experiments, not reported here, indicated

| Years | 1.2.1 (11) | 1.4.1 (2269) | 1.4.1 (2271) | 1.4.1 (2270) | 1.5.1 (2957) | 8.4.2 (1071) | 8.4.2 (1073) | 8.4.2 (1087) | 8.4.2 (1076) |
|---|---|---|---|---|---|---|---|---|---|
| 2000 | | 22.74099 | 31.29478 | 24.64515 | | 21000 | 6945939 | 1925978 | 200178 |
| ... | | | | | | | | | |
| 2005 | 33.7 | 27.47609 | 40.17754 | 30.41177 | | 243004 | 9645728 | 2188895 | 113076 |
| ... | | | | | | | | | |
| 2009 | 38.3 | 31.29914 | 48.66158 | 35.4571 | | 725012 | 9408891 | 2350763 | 113076 |
| ... | | | | | | | | | |
| 2015 | | 37.0523 | 62.26144 | 43.41761 | 17 | 1820623 | 12948523 | 2395294 | 107546 |
| ... | | | | | | | | | |
| 2018 | | | | | | | | | |

**Table 5.** Example time series for Afghanistan and a number of indicators arranged in columns

that for the imputation to have a chance of success a minimum of 25% of the values were required. In other words, given that our time series were of length 18, we needed values for five or more of the points. Thus, the filtering applied in Stage 2 also served to divide $TD$ in to two parts, $TD_{<5}$ and $TD_{\geq 5}$. Imputation could then be applied to $TD_{\geq 5}$.

A further issue with the collated time series is the different measures used with respect to the different indicators. For each country, as of February 2021, there were up to 3,408 different time series categories covering a wide range of domains[3]. For each of these time series one of 45 different units of measurement was used. Figure 6 lists each unit and the number of times it appeared in the data (up to February 2021). The dominant measuring unit is the percentage, followed by Tonnes and number. The percentage unit is widely used in the SDG data, as it is applicable to many different scenarios. The Tonnes unit of measurement is used most frequently with respect to Goal 8 "*Promote sustained, inclusive and sustainable economic growth, full and productive employment decent work for all*" where many sub-indicators measure material consumed in a country. The number unit measurement was often used to describe a monetary figure or a population. With this in mind, any consideration on building multivariate time series with the help of causal inference will require all the data to be on the same scale. Without scaling the time series, a series of population counts or of monetary value will always dominate over (say) series comprised of percentages values. There was thus also a requirement for some from of scaling to be applied to the data.

A set of experiments was conducted to find the best mechanism for imputing missing values and for scaling the data in $TD$. For the experimentation data for 41 countries was used together with three different imputation methods and three different scaling algorithms were considered. The imputation method were:

[3] Note that all sub-indicators are not necessarily relevant to all countries, for example sub-indicators concerned with forestation will not be relevant to a desert country, hence all countries do not feature exactly the same number of time series.

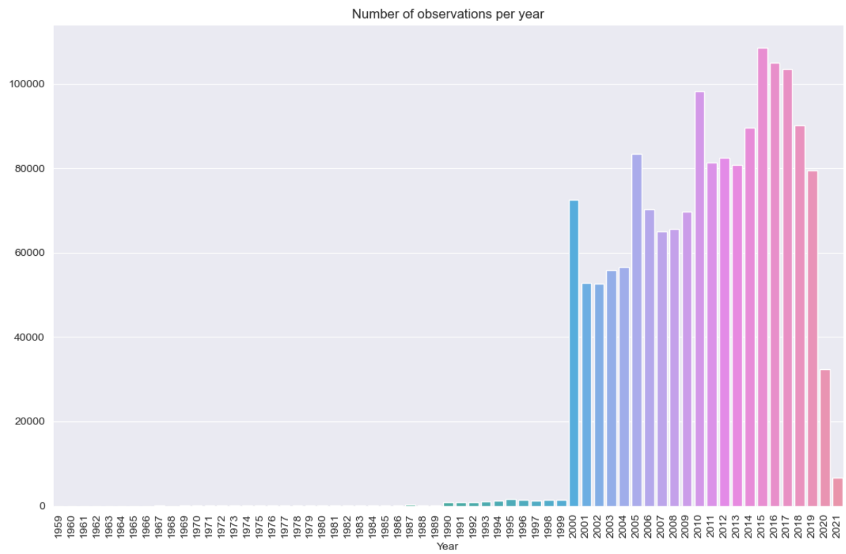


**Fig. 4.** Number of Observation per year with respect to the 30 different countries considered for evaluation purposes in this paper

(i) Spline, (ii) Time and (iii) Linear [11,15,29]. The scaling algorithms considered were: (i) Robust, (ii) Minmax and (iii) Standard [19]. These methods and algorithms were chosen because of their popularity in the literature. The three imputation methods and three scaling algorithms could be combined in nine different ways. Only complete time series where used for the experimentation; time series where all 18 values were available. There were 218 of these out of the 36,742 time series associated with the 30 countries used as a focus for the experimentation. Each time series was split into a training part and a testing part, $T_{i_{train}}$ and $T_{i_{test}}$, four values were then removed, at random, from each $T_{i_{train}}$ and then the selected imputation method and scaling were algorithm applied to $T_{i_{train}}$. The result was then used to predict the values in the test part. The imputation method and scaling algorithm that was the closest prediction match would then be used in Stage 2 of the SDG data pre-processing. For the prediction FBProhpet [25] was used. The adopted evaluation metric was average Root Mean Square Error (RMSE) was used. Although normally a lower RMSE values means a better results, in this experiment the goal was to stay as close as possible to the baseline RMSE value. The results of the experiments are given in Figure 5 which shows the average RMSE per combination. From the table it can be seen that the Time imputation coupled with Robust scaling (Time&Roubust) produced an average RMSE of 0.8296 which is the closest average RMSE values to the original value of 0.8185. This was then the combination used in Stage 2 to address the missing value and many measurement units used problem.

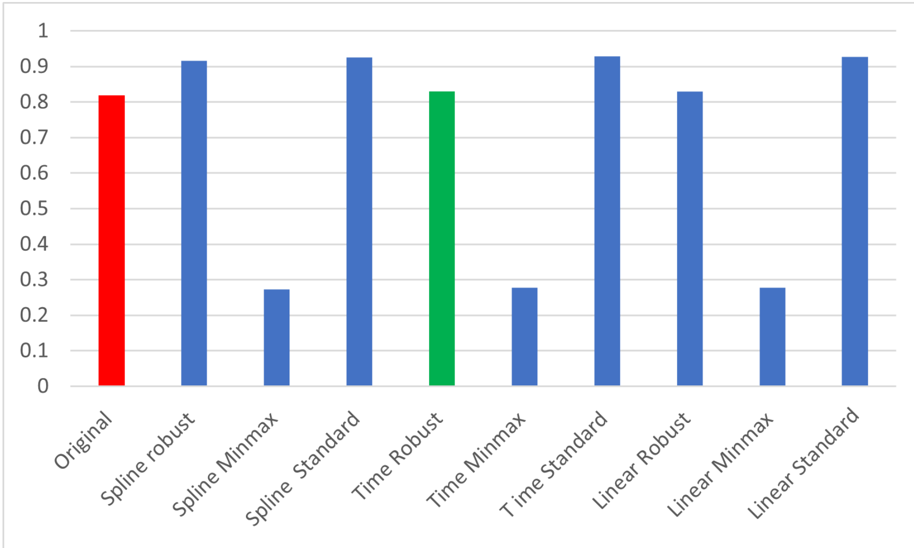


**Fig. 5.** comparison between the three imputation and three scaling methods considered

Returning to Figure 2, augmentation and scaling was applied to $TD_{\geq 5}$ to give $TD'_{\geq 5}$. However, note that $TD'_{<5}$ was not thrown away. The reason for the later will become clear later in this paper.

## 4 The Extended SDG Track, Trace and Forecast (SDG-TTF) Model

This section presents the SDG-TTF framework. The workflow for the framework is presented in Figure 7. The input is the set of time series, $\mathbf{T} = \{T_1, T_2, \ldots\}$ extracted from $TD'_{\geq 5}$ (generated as described in the previous section).

From the figure it can be seen that the extended SDG-TTF framework comprises five processes: (i) Data Grouping, (ii) Relation Discovery, (iii) multivariate Forecasting, (iv) univariate forecasting and (v) bottom-up classification. Note that two forecasting processes, multivariate and univariate, feed into the bottom up classification. The end result is a set of probabilistic SDG attainment predictions for the input set of countries. Each of the five stages will be considered in further detail in the remainder of this section.

During the data grouping process $\mathbf{T}$ is grouped into geographic regions. Recall that the objective of this paper is to improve on current SDG prediction effectiveness by taking into consideration both intra- and inter-causalities, causalities within individual countries and causalities between countries and their neighbours. Something not considered in previous work. The data grouping was conducted using geographic area codes based on the UN regional segmentation[4].

[4] https://unstats.un.org/sdgs/report/2019/regional-groups/

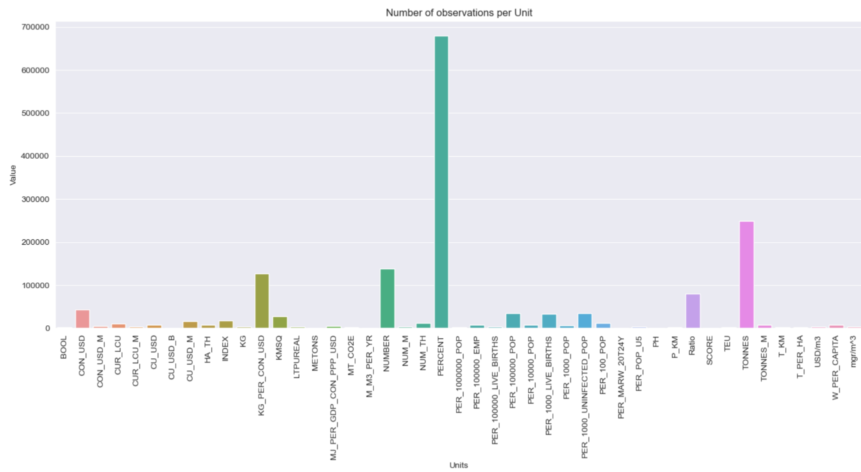


**Fig. 6.** Number of observations per unit

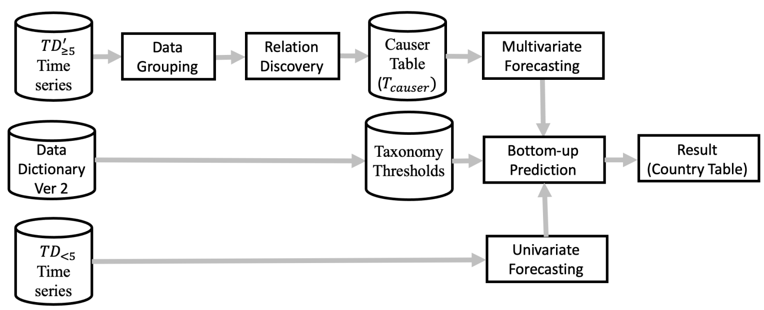


**Fig. 7.** Overview of the SDG-TTF workflow

The intuition here was that the SDG performance of neighbouring countries will have an impact on the SDG performance of the country under consideration. At the end of the grouping stage the set $\mathbf{T}$ will have been divide into a set of Region Time series, $\mathbf{T} = \{TR_1, TR_2, \dots\}$

The next process is to determine the causal relationships between the time series in each region time series set $TR_i$ for each grouping. Each $T_i \in TR_i$ is compared to each other time series in $TR_i$; in other words the complement of $T_i \in TR_i$ denoted as $T_i^c$. The interaction between each time series is measured using a causality ranking measure $r$. This is calculated using Root Mean Square Error (RMSE) in a similar manner as described for the experiments described in Sub-section 3.2. For the evaluation presented in the following section the six time series causality mechanisms listed in Section 2 were used (Lasso, $R^2$, Pearson Correlation, Mann-Whitney U Test, Granger and ACA). For each $T_i$, the time series in $T_i^c$ were ranked according to $r$ and the top $k$ selected as having a causal

| Year | 11.5.2<br>3189 | 9.c.1<br>2712 |
|---|---|---|
| 2000 | | 0 |
| 2001 | | 0 |
| 2002 | | 0 |
| 2003 | 0 | 0 |
| 2004 | 0 | 0 |
| 2005 | 0 | 0 |
| 2006 | 0 | 0 |
| 2007 | 0 | 0 |
| 2008 | 0 | 0 |
| 2009 | 0 | 0 |
| 2010 | 0 | 0 |
| 2011 | 0 | 0 |
| 2012 | 0 | 0 |
| 2013 | 0 | 0 |
| 2014 | 0 | 61 |
| 2015 | 0 | 89 |
| 2016 | 311 | 0 |
| 2017 | 0 | 95.1 |

**Table 6.** Two example time series that feature MAR and NMAR data for the geographic region of Egypt; 11.5.2 (3189), "*Direct economic loss attributed to disasters (current United States dollars)*" and 9.c.1 (2712), "*Proportion of population covered by a mobile network, by technology*"

relationship with $T_i$, the set $T^c_{i_k}$. For the evaluation presented later in this paper $k = 50$ was used. Each $T_i$ and $T^c_{i_k}$ pairing was then stored in a "causer table", $T_{causer} = \{\tau_1, \tau_2, \ldots\}$, where $\tau_i = T_i \cup T^c_{i_k}$.

For each $\tau_i \in T_{causer}$ the next process in the workflow, shown Figure 7, was to build a multi-variate time series forecasting model. Recall that these forecasting models sit at the leaf nodes of the SDG taxonomy populated with respect to a particular country. A range of tools and techniques are available whereby such a model can be constructed. However, for the evaluation presented later in this paper a multi-variate LSTM-Encoder-Decoder (Enc-Dec) [14] was used.

Recall, from the previous section, that during data preprocessing time series which were deemed unusable with respect to the determination of causality relationships were set aside in the set $T_{<5} = \{T_1, T_2, \ldots\}$. However, although unsuited to causality relationship determination this data can still be used for the purpose of forecasting SDG attainment. For each time series $T_i \in T_{<5e}$ a uni-variate time series forecasting model was built (to be held at the relevant leaf nodes within the taxonomy). Again there are a number of tools and techniques available whereby such a model can be constructed. For the evaluation presented in the following section uni-variate FBProphet was used.

The final process in the extended SDG-TTF workflow (Figure 7) is the prediction process where we ascertain whether a given country will meet its SDG goals or not using the generated multi-variate and uni-variate time series forecasting models described above. The fundamental process is similar to that of the SDG-AP framework presented in [2], which in turn was founded on the same hierarchical SDG topology described in [4] and used again with respect to the proposed extended SDG-TTF framework described here. The forecasting models are used to make predictions for individual indicators (leaf nodes in the topology) which are then compared to the threshold values held in the Data Dictionary generated as described in the previous section. The results are passed up the SDG topology hierarchy up to the root node. At each intermediate node a Boolean "and" operation will be applied and the result passed up the tree. The final result, that the given country will or will not attain its SDGs, will culminate at the root node. The results are stored in a "country table" and can be visualised using D3.js [5]. An example of the latter is given and discussed in Section 6 (Figure 8).

## 5 Evaluation

The evaluation of the proposed extended SDG-TTF model is presented in this section. For the evaluation the UN North Africa, South Asia and Northern Europe geographic regions, as defined by the UN Geoscheme, were considered:

**North Africa:** Algeria, Egypt, Libya, Morocco, Sudan, Tunisia and Western Sahara.

**South Asia:** Afghanistan, Bangladesh, Bhutan, India, Iran, Maldives, Nepal, Pakistan and Sri Lanka.

**Northern Europe:** Aland Islands, Denmark, Estonia, Faroe Islands, Finland, Iceland, Ireland, Isle of Man, Latvia, Lithuania, Norway, Svalbard and Jan Mayen Islands, Sweden and United Kingdom.

This comprised a total of 23,068 time series (leaf nodes in the topology), covering the 17 SDGs with respect to the sub-region of interest. After pre-processing (see Section 3) $TD'_{\geq 5}$ comprised 8,629 and $TD_{<5}$ 15,439 time series. The substantial number of time series allocated to $TD_{<5}$ was due to the large number of missing values that featured in the data.

The objectives of the evaluation were:

1. To determine the most appropriate causality discovery mechanism for use with the SDG-TFF framework.
2. To determine whether, by taking into consideration both intra-region and inter-region causality relationships, better SDG predictions could be produced.

For the evaluation the each time series was divided into two arts, the first 14 observations were used for training, and the last 4 observations for testing; $k = 50$ was used through out. All experiments were run on a windows 10 machine running under Ryzen 9 CPU, RTX 2060 GPU, 40 GB of RAM and 1TB SSD.

Comparisons were made with the SDG-AP and SDG-CAP prediction frameworks presented in [2] and [4] respectively. Recall that using SDG-CAP only intra-entity (single country) causal relationships were considered, as opposed inter-entity causal relationships as in the case of SDG-TTF. For SDG-AP framework two prediction models were considered, LSTM and FBProphet. All algorithms were implemented using the Python programming language. The evaluation metric used was RMSE (Root Mean Squared Error). As noted earlier, six different causality discovery mechanisms were considered: Lasso, $R^2$, Pearson Correlation, Mann-Whitney U Test, Granger and ACA.

| Time Series Code | | SDG-TTF | | | | | | SDG-CAP | SDG-AP | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Lasso | R2 | pearson | T_test | Granger causality | ACA | ACA | LSTM | FBProphet |
| 1 | afghanistan_1.4.1_2269 | 0.645 | 0.544 | 0.589 | 1.000 | 0.5585 | 0.6028 | 0.6910 | 0.313 | 0.0008 |
| 2 | afghanistan_1.4.1_2270 | 0.763 | 0.672 | 0.664 | 0.592 | 1.0000 | 0.6253 | 0.6337 | 0.310 | 0.0086 |
| 3 | afghanistan_1.4.1_2271 | 0.664 | 0.690 | 1.000 | 0.692 | 0.6158 | 0.7452 | 0.4122 | 0.297 | 0.0159 |
| 4 | afghanistan_1.4.1_2272 | 1.000 | 0.675 | 0.662 | 0.680 | 0.6749 | 0.6022 | 0.5578 | 0.308 | 0.0150 |
| 5 | afghanistan_1.4.1_2273 | 0.633 | 1.000 | 0.596 | 0.697 | 0.5002 | 0.5771 | 0.7197 | 0.288 | 0.0200 |
| 6 | afghanistan_1.4.1_2274 | 0.633 | 0.646 | 0.668 | 0.738 | 1.0000 | 0.6796 | 0.6149 | 0.312 | 0.0212 |
| 7 | afghanistan_1.a.1_2956 | 0.191 | 0.259 | 0.268 | 1.000 | 0.2522 | 0.2416 | 0.9103 | 0.985 | 0.5219 |
| 8 | afghanistan_1.a.2_2277 | 0.874 | 0.868 | 1.000 | 0.906 | 0.8732 | 0.9049 | 0.9575 | 3.304 | 1.1733 |
| 9 | afghanistan_10.4.1_2721 | 1.000 | 0.296 | 0.308 | 0.300 | 0.3159 | 0.3216 | 0.3675 | 1.308 | 0.0725 |
| 10 | afghanistan_10.5.1_2725 | 0.251 | 0.277 | 1.000 | 0.261 | 0.2649 | 0.2505 | 0.2716 | 0.298 | 2.0310 |
| 11 | afghanistan_10.5.1_2726 | 0.047 | 1.000 | 0.016 | 0.016 | 0.0136 | 0.0239 | 0.0179 | 0.283 | 3.5772 |
| 12 | afghanistan_10.5.1_2727 | 1.000 | 0.007 | 0.056 | 0.130 | 0.0575 | 0.0854 | 0.1129 | 0.160 | 5.8625 |
| Average | | 0.642 | 0.578 | 0.569 | 0.584 | 0.510 | 0.471 | 0.522 | 0.680 | 1.110 |
| Standard Deviation | | 0.311 | 0.299 | 0.330 | 0.318 | 0.325 | 0.265 | 0.281 | 0.856 | 1.7833 |

**Table 7.** A sample of RMSE values for selected SDG indicators for Afghanistan

A sample of the the recorded RMSE values for the country Afghanistan and 12 selected SDGs is given in Tables 7. The first two columns give the time series ID number and the Individual Series (IS) indicator. The next six columns give the RMSE results obtained using the six causality mechanisms considered and the extended SDG-TTF framework. The following column gives the results obtained using SDG-CAP and ACA causality as described in [4] and the last two columns the results obtained using SDG-AP coupled with Univariate LSTM and FBProphet as described in [2]. The average RMSE value is given at the bottom of the table, for each approach considered, together with the associated standard deviation. From the sample it can be seen that the Extended SDG-TTF framework, coupled with ACA causality, produced the best overall result (highlighted in bold font). These results were confirmed by inspection of the complete set of results (not shown here) for the three regions considered. It was thus concluded that the most appropriate causality discovery mechanism was the ACA mechanism.

Tables 8, 9 and 10 present a summary of the results obtained for the North Africa, South Asia and Northern Europe regions considered, suing: SDG-TTG and ACA causality, SDG-CAP and ACA causality, SDG-AP with LSTM and SDG-AP with FBProphet. From tables it is clear that consideration of inter-entity causal relationships, as well as intra-entity causal relationships, as incorporated into the SDG-TTF framework, results in an improved SDG attainment prediction.

| **Country** | **SDG-TTF (ACA)** | | **SDG-CAP (ACA)** | | **SDG-AP (FBProphet)** | |
|---|---|---|---|---|---|---|
| **RMSE** | AVG | SD | AVG | SD | AVG | SD |
| Algeria | 0.3 | 0.5 | 0.4 | 0.9 | 0.8 | 7.6 |
| Egypt | 0.4 | 1.4 | 0.5 | 2.0 | 0.6 | 3.1 |
| Libya | 0.8 | 1.1 | 0.9 | 1.0 | 0.6 | 0.8 |
| Morocco | 0.6 | 0.3 | 0.5 | 1.4 | 0.6 | 1.3 |
| Sudan | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 | 0.4 |
| Tunisia | 0.4 | 0.8 | 0.5 | 1.1 | 0.7 | 1.8 |
| Western Sahara | 0.5 | 0.3 | 0.6 | 0.5 | 0.8 | 0.5 |
| **Average** | 0.4 | 0.7 | 0.5 | 1.0 | 0.6 | 2.2 |

**Table 8.** Average RMSE values for the North Africa geographic region per country [3]

## 6 System Operation

The operation of the SDG-TTF framework was investigated using a number of case studies. One such case study is partly presented here. Namely, SDG 3, Target 2: "*By 2030, end preventable deaths of newborns and children under five years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1000 live births and under-5 mortality to at least as low as 25 per 1000 live births*", and the country Algeria. Target 3.2 thus comprises two indicators (3.2.1 and 3.2.2), "*Under-five mortality rate*" and "*Neonatal mortality rate*". Note that neonatal interpreted as aged less over 1 month old. For the first we have four sub-indicators (time series): (i) deaths of female children aged less than one (1Y/F), (ii) deaths of male children aged less than one (1Y/M), (iii)

| **Country** | **SDG-TTF (ACA)** | | **SDG-CAP (ACA)** | | **SDG-AP (FBProphet)** | |
|---|---|---|---|---|---|---|
| **RMSE** | AVG | SD | AVG | SD | AVG | SD |
| Afghanistan | 1.305 | 7.228 | 1.310 | 7.231 | 1.887 | 2.424 |
| Bangladesh | 1.005 | 3.751 | 1.089 | 2.614 | 1.780 | 2.611 |
| Bhutan | 1.481 | 8.055 | 1.452 | 4.354 | 3.352 | 40.989 |
| India | 1.480 | 8.079 | 0.744 | 0.812 | 0.984 | 2.481 |
| Iran | 1.321 | 7.666 | 0.760 | 1.194 | 1.938 | 19.521 |
| Maldives | 1.298 | 1.285 | 3.624 | 23.232 | 0.991 | 3.256 |
| Nepal | 7.658 | 7.230 | 1.267 | 1.643 | 1.518 | 3.844 |
| Pakistan | 0.490 | 0.507 | 0.930 | 1.672 | 3.754 | 65.313 |
| Sri Lanka | 0.436 | 0.520 | 1.181 | 2.230 | 1.958 | 2.964 |

**Table 9.** Average RMSE values for the South Asia geographic region per country [3]

deaths of female children aged less than five (5Y/F) and (iv) deaths of male children aged less than five (5Y/M). The threshold in this case is $\leq 25$ *per* 1000. For the second we have one sub-indicator (time series): deaths of children under one month of age (1Month/FM) for which the threshold is $\leq 12$ *per* 1000.

SDG-TTF was then used to make predictions for the year 2030. The generated output is a "country table", as indicated in the workflow presented in Figure 7. A fragment of this table for Target 3.2 is given in Table 11. The first column gives the GTI. The second gives the sub-indicator (as described above). The third gives the gives the mortality value per 1000 live births in 2015 which is the base year for SDGs. The fourth gives the target thresholds for TGI 3.2.1 and 3.2.2, $\leq 25$ *per* 1000 and $\leq 12$ *per* 1000 respectively. The fifth column gives the predicted mortality value per 1000 live births for 2030. The sixth column gives the binary classification, "Met" or Not Met". For Target 3.2 to be attained (met), the predicted value for each sub-indicator must meet its threshold (at or below the relevant threshold value in this case). In this partial example, all of the included sub-indicators for indicators GTI 3.2.1 are met, unfortunately GTI 3.2.2 is not met.

The software for the Extended SDG-TTF framework includes a visualisation mechanism, as indicated in Figure 7. This was implemented using D3.js [5]. The visualisation allows users to: (i) track the progress of different goals over a given time frame, and (ii) trace the achievement of individual bottom level indicators in an interactive manner. An example of such a visualisation is given in Figure 8 using the case study presented above. From the figure it can be seen that using the visualisation it is easy to identify goal attainment (or non-attainment as in this case). Nodes coloured in green highlight goals, targets, indicators and sub-indicators that will be attained on time. Nodes coloured in red highlight goals, targets, indicators and sub-indicators that will not be attained on time. For a more detailed analysis of why a goal is not attaining the relevant country table can be inspected.

| **Country** | **SDG-TTF (ACA)** | | **SDG-CAP (ACA)** | | **SDG-AP (FBProphet)** | |
|---|---|---|---|---|---|---|
| **RMSE** | AVG | SD | AVG | SD | AVG | SD |
| Aland Islands | 0.187 | 0.125 | 0.197 | 0.149 | 0.227 | 0.189 |
| Denmark | 0.637 | 0.724 | 4.421 | 70.753 | 2.249 | 24.416 |
| Estonia | 0.644 | 0.726 | 1.705 | 3.792 | 3.567 | 31.441 |
| Faroe Islands | 0.647 | 0.724 | 0.877 | 1.100 | 0.381 | 0.693 |
| Finland | 0.644 | 0.722 | 1.580 | 3.562 | 0.813 | 4.094 |
| Iceland | 0.644 | 0.728 | 0.360 | 0.463 | 23.875 | 317.859 |
| Ireland | 0.608 | 0.711 | 0.682 | 0.158 | 261.329 | 4989.750 |
| Isle of Man | 0.607 | 0.707 | 0.572 | 0.480 | 0.156 | 0.316 |
| Latvia | 0.348 | 0.303 | 0.592 | 0.352 | 1.010 | 3.822 |
| Lithuania | 0.348 | 0.444 | 0.585 | 0.304 | 26.780 | 486.870 |
| Norway | 1.675 | 1.537 | 0.219 | 0.272 | 0.728 | 2.488 |
| Svalbard and Jan Mayen Islands | 0.401 | 0.0274 | 0.388 | 0.248 | .0708 | 0.577 |
| Sweden | 0.313 | 0.355 | 0.398 | 0.479 | 2.574 | 15.406 |
| United Kingdom | 0.336 | 0.355 | 1.080 | 1.714 | 1799.020 | 37049.630 |

**Table 10.** Average RMSE values for the Northern Europe geographic region per country

| **GTI** | **Age/Sex** | **Initial** | **Target** | **Forecast** | **Result** |
|---|---|---|---|---|---|
| 3.2.1 | 1Y/F | 20.2 | <=25 | 16.94 | Met |
| 3.2.1 | 1Y/M | 22.9 | <=25 | 20.82 | Met |
| 3.2.1 | 5Y/F | 23.7 | <=25 | 19.89 | Met |
| 3.2.1 | 5Y/M | 26.6 | <=25 | 24.13 | Met |
| 3.2.2 | 1Month/FM | 15 | <=12 | 13.75 | Not Met |

**Table 11.** Forecast results for Target 3.2, for the target year 2030, and the country Algeria [3]

## 7 Conclusion

In this paper, we have presented an extended analysis of the SDG-TTF attainment prediction framework [3], which, unlike previous frameworks directed at SDG attainment prediction, considers inter- and intra-geographic entity (county, region) causal relationships. It is argued in this paper that individual SDG subindicators should not be considered in isolation, in other words in terms of an individual time series, because inspection of the indicators demonstrates clear potential for causal relations with other indicators for a given geographic entity, and potential for causal relationships with the indicators for neighbouring geographic region. The evaluation of the framework shows that more accurate SDG attainment predictions using the SDG-TTF framework can be made. For future work, the authors intend to expand the investigation using more data sources than simply the SDGs data, and consider using alternative causal relationship discovery mechanisms. Finally, the authors intend to evaluate the effect of nat-

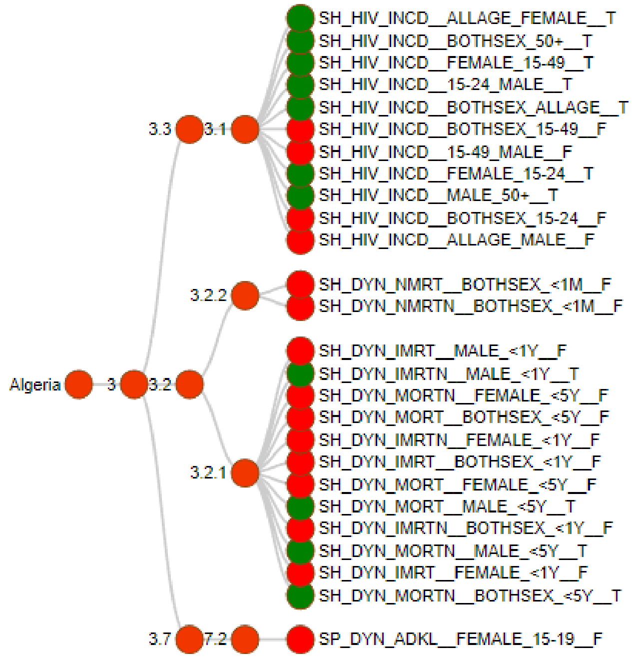


**Fig. 8.** Visualising of SDG attainment for part of goal 3 [3]

ural disasters, such as the COVID-19 pandemic, which occur for short periods, on SDG attainment prediction. Note that all the data provided in this paper can be found in the project Github reposetery[5].

## References

1. Alam, N., Rudin, C.: Robust Nonparametric Testing for Causal Inference in Observational Studies. Optimization Online, Dec pp. 1–39 (2015)
2. Alharbi, Y., Arribas-Bel, D., Coenen, F.: Sustainable development goal attainment prediction: A Hierarchical Framework using Time Series Modelling. IC3K 2019 **1**, 297–304 (2019). https://doi.org/10.5220/0008067202970304
3. Alharbi, Y., Arribas-Bel, D., Coenen, F.: Sustainable development goals monitoring and forecasting using time series analysis. In: Fred, A.L.N., Sansone, C., Madani, K. (eds.) Proceedings of the 2nd International Conference on Deep Learning Theory and Applications, DeLTA 2021, Online Streaming, July 7-9, 2021. pp. 123–131. SCITEPRESS (2021). https://doi.org/10.5220/0010546101230131, https://doi.org/10.5220/0010546101230131
4. Alharbi, Y., Coenen, F., Arribas-Bel, D.: Sustainable development goal relational modelling: Introducing the sdg-cap methodology. In: DAWAK. vol. 12393 LNCS, pp. 183–196 (2020). https://doi.org/10.1007/978-3-030-59065-9_15

[5] https://github.com/Yassir-Alharbi/Sustainable-Development-goals.

5. Bostock, M., Ogievetsky, V., Heer, J.: D3 data-driven documents. IEEE TVCG **17** (dec 2011). https://doi.org/10.1109/TVCG.2011.185, http://dx.doi.org/10.1109/TVCG.2011.185
6. Chen, K., Zhou, Y., Dai, F.: A LSTM-based method for stock returns prediction: A case study of China stock market. In: IEEE Big Data 2015. IEEE (2015). https://doi.org/10.1109/BigData.2015.7364089
7. De Gooijer, J.G., Hyndman, R.J.: 25 Years of Time Series Forecasting. IFJ **22**(3), 443–473 (2006). https://doi.org/10.1016/j.ijforecast.2006.01.001, http://www.sciencedirect.com/science/article/pii/S0169207006000021
8. Dörgo, G., Sebestyén, V., Abonyi, J.: Evaluating the interconnectedness of the sustainable development goals based on the causality analysis of sustainability indicators. Sustainability (Switzerland) **10**(10), 3766 (2018). https://doi.org/10.3390/su10103766
9. Epprecht, C., Guegan, D., Veiga, Á.: Comparing variable selection techniques for linear regression: LASSO and Autometrics. Centre d'économie de la Sorbonne (2013), http://halshs.archives-ouvertes.fr/halshs-00917797/
10. Frey, B.B.: Pearson Correlation Coefficient. In: The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation, pp. 1–4. Springer (2018). https://doi.org/10.4135/9781506326139.n510
11. Hall, C.A., Meyer, W.W.: Optimal error bounds for cubic spline interpolation. Journal of Approximation Theory **16**(2), 105–122 (1976). https://doi.org/10.1016/0021-9045(76)90040-X
12. Heitjan, D.F., Basu, S.: Distinguishing "missing at random" and "missing completely at random". The American Statistician **50**(3), 207–213 (1996)
13. Hyndman, R., Kostenko, A.: Minimum sample size requirements for seasonal forecasting models. Foresight **6**(Spring), 12–15 (2007)
14. Jason, B.: Deep Learning For Time Series Forecasting, vol. 1. Machine Learning Mastery (2018)
15. Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J., Kolehmainen, M.: Methods for imputation of missing values in air quality data sets. Atmospheric Environment **38**(18), 2895–2907 (2004). https://doi.org/10.1016/j.atmosenv.2004.02.026
16. Lozano, C.J.: Measuring progress from 1990 to 2017 and projecting attainment to 2030 of the health-related Sustainable Development Goals for 195 countries and territories: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet **392**(10159), 2091–2138 (2018). https://doi.org/10.1016/S0140-6736(18)32281-5
17. Narayan, P.K., Smyth, R.: Multivariate granger causality between electricity consumption, exports and GDP: Evidence from a panel of Middle Eastern countries. Energy Policy **37**(1), 229–236 (2009). https://doi.org/10.1016/j.enpol.2008.08.020
18. Nauta, M., Bucur, D., Seifert, C.: Causal discovery with attention-based convolutional neural networks. Machine Learning and Knowledge Extraction **1**(1). https://doi.org/10.3390/make1010019
19. Pedregosa: Scikit-learn: Machine learning in python. Journal of Machine Learning Research **12** (2011)
20. Qing, X., Niu, Y.: Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy . https://doi.org/10.1016/j.energy.2018.01.177
21. R González, L., B, D., M, L.F., V, R.: Long-term electricity supply and demand forecast (2018-2040): A LEAP model application towards a sustainable power generation system in Ecuador. Sustainability (Switzerland) **11**(19), 5316 (2019). https://doi.org/10.3390/su11195316

22. Rockström, J.: Johan rockström and pavan sukhdev present new way of viewing the sustainable development goals and how they are all linked to food.: Stockholm resilience centre (2016)
23. Seabold, S., Perktold, J.: Statsmodels: Econometric and Statistical Modeling with Python. In: the 9th Python in Science Conference. pp. 92–96 (2010). https://doi.org/10.25080/majora-92bf1922-011
24. T, A., L, N., S, G., Ps, D., K, M.: Efficiency forecasting for municipal solid waste recycling in the context on sustainable development of economy. In: E3S Web of Conferences. vol. 166, p. 13021 (2020). https://doi.org/10.1051/e3sconf/202016613021
25. Taylor, S.J., Letham, B.: Forecasting at Scale. American Statistician **72**(1), 37–45 (2018). https://doi.org/10.1080/00031305.2017.1380080
26. Tibshirani, R.: Regression Shrinkage and Selection Via the Lasso. JRSS (1996). https://doi.org/10.1111/j.2517-6161.1996.tb02080.x, http://www.jstor.org/stable/2346178
27. UN, S.D.: E-Handbook on Sustainable Development Goals Indicators
28. United Nations: The Millennium Development Goals Report. United Nations p. 72 (2015). https://doi.org/978-92-1-101320-7
29. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods **17**, 261–272 (2020). https://doi.org/10.1038/s41592-019-0686-2