A concept of forecasting origin-destination air passenger demand between global city pairs using future socio-economic scenarios

Ivan Terekhov[*], Robin Ghosh[†] and Volker Gollnick[‡]

German Aerospace Center (DLR), Air Transportation Systems, Blohmstr. 18, Hamburg, 21079, Germany

This study introduces a concept of a new method of forecasting air passenger flows on a global level using socio-economic scenarios. The method has two steps: forecasting the topology of origin-destination demand network and predicting the number of passengers on existing and new connections. Network theory is applied to simulate demand connections between cities utilizing weighted similarity based algorithms. The number of passengers on a connection is defined using quantitative analogies. Preliminary calculations show promising results. This concept of the global passenger demand prediction will be applied in a modular environment modeling the future air transport system.

I. Introduction

Air transport system (ATS) is a multi-disciplinary complex system with various interactions between stakeholders within the system and its environment. The "atomic model" of the ATS in Fig.1, introduced by Ghosh[1], outlines the core stakeholders in the system, united by the aircraft as the major connecting element and external elements. Changes in any element of this complex interconnected system may cause changes in the whole system.

The ATS growth has a clear correlation to world economy growth. The number of worldwide passengers within the ATS has increased from 1.5 to 3 billion from 2002 to 2013 (based on the ADI database[§]). Such growth has undoubtedly an increasing affect on the environment. Airline schedules, network and fleet decisions are based on a number of drivers, but arguably the most important is the available origin destination city-pair passenger demand. It is likely that there will be a number of cities with significant air traffic connections that have no air traffic connections today. This is particularly true for countries such as China. By not including air traffic to these growing cities, global growth in air traffic and emissions would be underestimated, resulting in the corresponding underestimate of the climate impacts associated with aviation. Therefore, there is the particular importance to capture the full environmental impact of future growth in the aviation sector. As shown by Lee[2], air transport contributes 2-3% of global CO_2 emissions and 3.5-4.9% of global radiative forcing, if non-CO_2 effects are included. Today, the impacts of CO_2 emissions on the environment are closely studied. However, non-CO_2 effects have not been subject to the same level of study. Accordingly, a more robust scientific understanding of the effects of non-CO_2 emissions is still needed[3]. The non-CO_2 emissions have different impacts on the environment in different regions of the world. For example, NOx has been shown to induce short-lived greenhouse gas ozone. The gas produced at the equator has a higher radiative forcing than the same amount of emissions in northern regions. This implies that geographical information of a flight route, such as location of departure and destination airports, as well as flight path, is essential for assessing the impact of non-CO_2 emissions. Accordingly, to assess the non-CO_2 impact, the number of flights and type of aircraft operated on routes must be known so as to quantify the amount of such emissions on a global scale. To obtain this information the number of passengers on these routes must be estimated. Finally, to make such estimations, air passenger demand between origin-destination has to be determined.

[*] PhD candidate, Department System Analysis Air Transport.
[†] Research Engineer, Department System Analysis Air Transport.
[‡] Head of the DLR Air Transportation Systems.
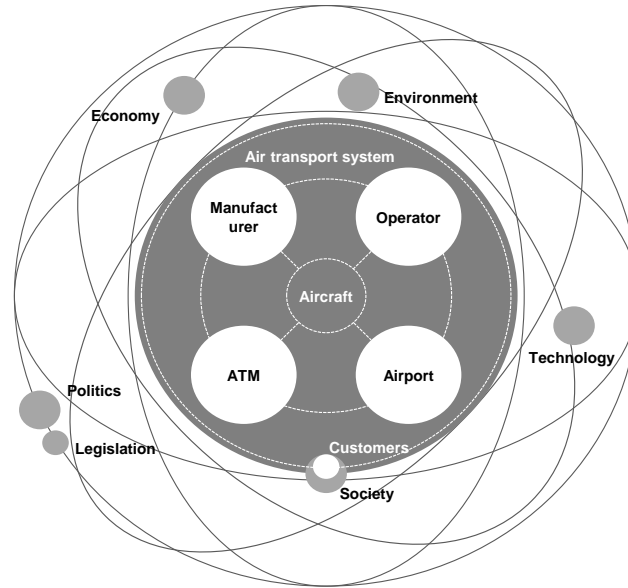[§] Sabre Aviation Data Intelligence

Figure 1: ATS model

This assessment approach of non-CO_2 emissions takes into account different layers of the ATS starting from origin-destination air passenger demand to trajectories of an aircraft. Thus, there is a need for a methodology that is capable to describe the interactions between elements and changes in the ATS. The German Aerospace Center (DLR) project known as "*WeCare*"** studies the potential of climate efficient flight by using forecast weather information on a global scale up to a 2050 time horizon, in the context of the ATS. One of the *WeCare* project's aim is to assess the impact of non-CO2 emissions on a global scale. Within this project, DLR Air Transportation Systems is developing a modular environment known as "*AIRCAST*"[4] (air travel forecast), which aims to forecast future development of the ATS based on socio-economic scenarios. *AIRCAST* allows DLR to simulate a range of possible outcomes for the future development of the ATS and assess, for example, the impact of new technology on the number of demand passengers or the size and number of aircraft on particular routes. From this, a chain of models for the future ATS has been developed (Fig.2). The demand forecast model of 'origin-destination air travel passenger demand between city-pairs' called "*D-CAST*", based on socio-economic scenario, is the first layer in a
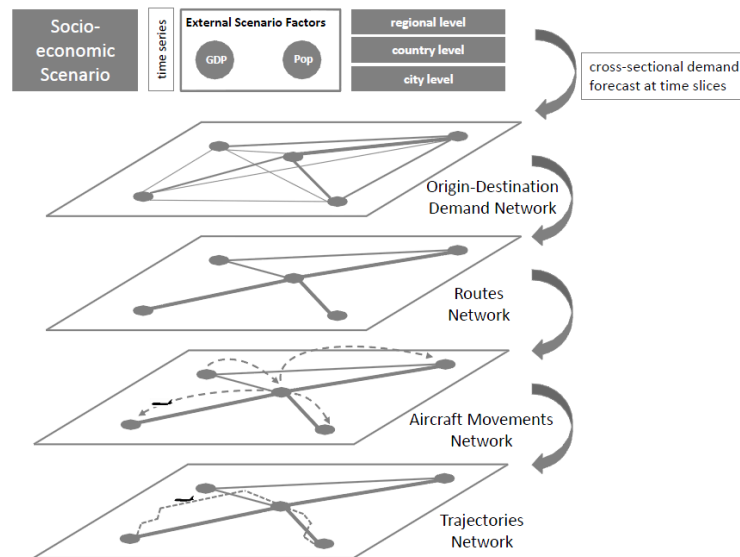

Figure 2: 4-layers approach

** *WeCare* - utilizing **we**ather information for **c**lim**a**te efficient and eco efficient futu**re** aviation.

American Institute of Aeronautics and Astronautics

chain of models within *"AIRCAST"*. Since one of the *WeCare* project's aim is to assess the impact of non-CO2 there is a high importance of geographical information of flights. The demand forecast model has to take into account not only the number of air passengers, but also the possibility of changes to the number origin-destination demand pairs. These pairs are formed by passengers traveling by air between an origin and a destination, regardless of any intermediate stops. To meet these requirements, the model has to include cities where at least one airport is present. The model must take into account these worldwide cities and simulate air passenger demand connections between them and the number of passengers on these connections within the forecast period.

This paper is organized as follows. Section 2 introduces the assumptions and definitions used in the study. Section 3 provides an overview of related studies in this area. Section 4 describes the data sources utilized for the study. Section 5 presents the concept for forecasting air passenger demand. Section 6 presents the preliminary results of the study. Section 7 summarizes the progress of the study to date and gives an outline of the future research.

II.   Demand terminology

Given the complexity and the novelty of this study, it is important that the assumptions and terminology used are clearly defined at the outset. *The first assumption* is that every person in the world has a latent demand to travel by air. *The second assumption* is that a particular set of individual conditions is required to give rise to a decision of a person to travel by air. In other words, if these certain personal conditions are met, a person *will* choose to travel by air. Such conditions could include socio-economic indicators (e.g. GDP, population, oil price, etc.) as well as ATS specific indicators (e.g. travel time, frequency, number of transfers, airfare, etc.). Based on these assumptions two categories can be defined: people whose conditions to travel by air are met constitute *the realized demand*; all other people whose conditions to travel by air are not met constitute *the unrealized demand*. Accordingly, the sum of realized and unrealized demand will make up the total demand for air travel. In other words, the sum of realized and unrealized demand is equal to the latent demand. According to the first assumption, the latent demand is the world population. For example, unrealized demand implies that a person does not travel by air due to various combinations of reasons (e.g. high ticket price, long journey time, etc.). In opposite, realized demand implies that a person does travel by air because individual and ATS-specific requirements of the person (e.g. ticket price, journey time, itinerary, etc.) are satisfied.

In addition, *directed air passenger demand* is the demand in direction from point A to point B, but does not include the demand in direction from point B to A. The sum of directed air passenger demand on directions A-B and B-A represent the *undirected air passenger demand* between points A and B. In other words, undirected demand is the total air passenger demand between two points regardless of directions.

III.   Literature review

Forecast of air passenger demand is an important basis for planning in the constantly changing aviation transportation system. The aircraft industry, and researchers, study air passenger demand and develop forecast models using various techniques and levels of aggregation.

A. Industry forecasts

As discussed in Doucet et al. paper[5], an origin destination air passenger demand model is an important part in the Airbus Global Market Forecast (GMF)[6] methodology. The GMF methodology for forecasting future ATS contains three basic steps: traffic forecast to the next 20 years, the network forecast[††] and a forecast of the number of required aircraft. Air passenger demand forecast in the GMF is a part of the second step. For the network forecast, initially, a traffic forecast between countries is disaggregated to a set of city pairs. Next, flight segments are modeled between any two cities in the set. The obtained flight segments network includes existing routes as well as future possible routes. Utilizing a market share model, a percentage of air passengers is assigned to each flight segment. Finally, the number of passengers is defined as the percentage of passengers on each flight multiplied by the origin-destination demand between cities. The origin-destination air passenger demand model utilizes a modified gravity model to forecast the number of passengers between 279 cities around the world. The modified gravity model takes into account a spatial dependence between origin and destination. In other words, the model takes into account the impact to air passengers flow between cities by utilizing characteristics at proximal cities. From the perspective of the present study, and its categorization of realized and unrealized demand, the air passenger demand model discussed in Ref. 4, is dealing with *realized directed* air passenger demand.

[††] Here „network forecast" implies a forecast of routes between cities

American Institute of Aeronautics and Astronautics

The United Kingdom Department for Transport's UK Aviation Forecasts 2013[7] includes the National Air Passenger Demand Model. This model uses a combination of a set of time series econometric models of past UK air passenger demand including projections of key driving variables and assumptions about how the relationship between UK air travel and its key drivers will change into the future. The model provides forecasts for domestic destinations within the UK, international regions of origin for flights into to the UK and international passengers connecting through UK airports.

Other industry forecasts are mainly predicting Revenue Passengers Kilometers (RPK). These forecasts do not present a separate air passengers demand model. The Boeing Current Market Outlook 2013-2032[8] forecast uses an empirical equation where RPK growth between regions is equal to sum of GDP growth and a time-varying function. The function is not directly associated with GDP growth. This component of growth derives from the value travelers place on the speed and convenience that only air travel can offer. In the Worldwide Market Forecast for Commercial Air Transport[9] the Marketing Japan Aircraft Development Corporation developed a traffic forecast that predicts RPK between 11 world regions. The relationship between past RPK, GDP and "Yield" is analyzed by each region to obtain their regression equation, however the equation is not provided in their publication.

B. Academic studies

The Aviation Integrated Modelling (AIM)[10,11] project was initiated by the University of Cambridge, UK. The aim of this project is to develop a tool to assess different current and future policies in aviation[12]. The AIM project contains a set of connected modules that were created to fulfill the policy assessment goals of the project. These modules are Aircraft Technology & Cost Module, Air Transport Demand Module, Airport Activity Model, Aircraft Movement Module, Global Climate Module, Local Air Quality & Noise Module, and Regional Economic Module. The Air Transport Demand Module deals with true origin-destination (OD) air passenger and freight demand. Currently, this module contains a simple model on city level which considers realized undirected air passenger demand. The model is represented as a gravity model with OD connections between 700 cities around the world. The main variables in the gravity equation are: average local per capita income, greater metropolitan area or equivalent population and generalized cost to a passenger of air travel between cities. The gravity equation has been adapted to short-haul, medium-haul and long-haul as well as for different regions. The equation has been calibrated on current and historical data.

Suryani et al.[13] model air passenger demand and passenger terminal capacity expansion using a systems dynamics approach. The study concentrates on single airport level. Their model predicts when an airport should expand runway capacity, passenger terminal capacity and to determine the total airport area needed to meet future demand. Alam and Karim[14] address the present condition of the air transportation system in Bangladesh. They analyze the operation and level of service of the system, realized undirected demand and supply structure and the network configuration. A stepwise multiple linear regression analysis, using a time series collected for five years, was utilized to calculate total passenger trips per week along existing routes. Grosche et al.[15] present gravity models for the estimation of air passenger volume between city-pairs. The estimation is based on socio-economic and geographic factors for the fixed number of city-pairs. Thus, this approach did not take into account the possibility of new city-pairs within the air transport system.

C. Discussion

The studies mentioned above have utilized a range of techniques and considered various levels of aggregation. Industry forecasts and academic studies show various methods to calculate the demand in particular airports[11], on particular routes[12, 13], on regional level[6] or on city level with fixed number of connections between cities[9, 10]. Mostly, the aforementioned forecasts deal with realized undirected air passenger demand using gravity models. Gravity models have to be calibrated for different types of city-pairs (e.g. short-haul, medium-haul, long-haul, international, regional, local, etc.). However, when dealing with larger numbers of city pairs, the complexity of the calibration requirements of these models increases. Moreover, these studies do not include a method of forecasting an evolution of air passenger demand between cities at a global level. They fail to take into account the potential for changes in the number of airport-connected cities when forecasting demand within an air transport system.

IV.   Data

In *AIRCAST,* 2012 has been adopted as the base year. For the base year, required data have been obtained from Sabre Airport Data Intelligence (ADI) database: origin-destination city pairs worldwide, air passenger demand and average airfare between these cities. Additionally, data for GDP, population and geographical coordinates of the cities have been obtained from various databases (see below).

The ADI database contains booking information from the Global Distribution System (GDS), its primary data source, and other external data sources[16]. In the ADI database it is possible to obtain information of passenger numbers between origin and destination airports as well as average airfare at OD level. The ADI database presents the realized demand for air travel. Due to the assumption that air passenger demand is generated on city level and not on airport level, the OD air passenger demand data from ADI has been aggregated to the city level. Thus, 4435 cities and 533170 connections between them have been obtained. In addition, for each city the GDP (utilizing UN[17] and World Bank[18] data), population (UN[19] and MaxMind[20] data) and geographical coordinates (OurAirports[21] and OpenFlights[22] data) have been retrieved. The final database for the base year 2012 includes OD city pairs, socio-economic indicators (city GDP, city population), geographical coordinates, great circle distances between OD cities and average airfare between these cities. Various socio-economic scenarios could be used as alternative inputs for forecasting air passenger demand. These scenarios contain annual data for GDP, population on city level and oil price in a given period of time.

V. Method

This paper presents a concept to forecast the evolution of the air travel passenger demand between cities based on socio-economic scenarios, taking into consideration the probability of changes to the number of the origin-destination demand connections within the ATS over time. In other words, the proposed concept forecasts passenger demand as well as topology changes of the 'air passenger demand network', within the forecast period. The method computes air passenger demand at any given point of time within the forecast period. Within this study the demand forecasting model considers realized undirected air passenger demand.

The method has two steps: forecasting the topology of the origin-destination demand network and calculating demand on existing and new connections. The first step of the method, determines whether the demand connection between a given city pair exists or not. This is done by implementing a weighted similarity-based algorithm. The weight is represented by a combination of socio-economic information of cities in pairs, and the distance between them. The second step of the method, based on the existence of air passenger demand between cities, seeks to forecast the realized air passenger demand between these cities.
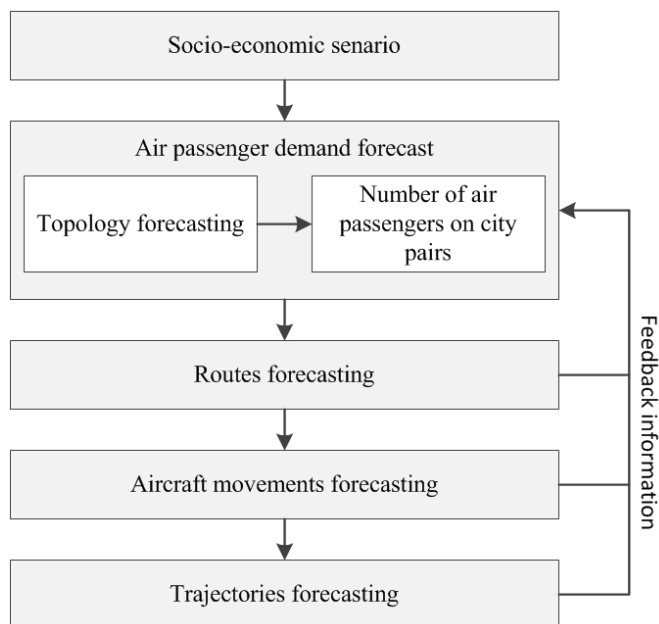

Figure 3: Set of models for forecasting the future ATS

The method computes realized air passenger demand using a quantitative analogies method that takes into account the socio-economic indicators of cities, the airfare information and geographical characteristics of given city-pairs, at a given forecast period and for the base year (where all the required information about cities and the connections is known). Next, the obtained number of passengers is calibrated by feedback information from routes, aircraft movements and trajectory levels (see Fig. 3).

For defining air passenger demand between cities in the demand network for one discrete "slice" of a socio-economic scenario, a two-stage method has been proposed: city-pairs definition and air passenger demand on city-pairs definition.

A. Topology forecast

The socio-economic characteristics of the cities change throughout the duration of the scenario. Therefore, the original demand network topology transforms as a result of new air passenger demand connections appearing and disappearing. These changes to the network topology must be taken into account when determining the air passenger demand between cities. Furthermore, the number of cities with airports will change throughout the duration of the scenario. A city is added (eliminated) to (from) the demand network when an airport in a city is appearing (disappearing) which did not have an airport before. A city is eliminated from the demand network when an airport in a city is disappearing so that it has no airport anymore. The basic structure of the topology forecast is shown at Fig. 4.
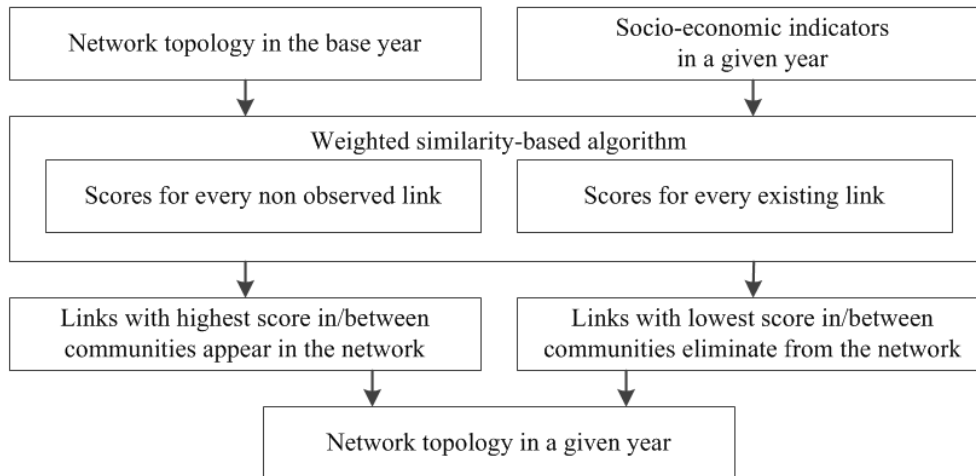

Figure 4: Topology forecast structure

Air passenger demand could be described as a network where cities are represented as nodes and city pairs as edges. Thus, the problem of forecasting the potential of OD demand between city-pairs turns to link prediction between nodes in a given year of a socio-economic scenario. Many studies of network theory have been dedicated to link prediction, and these methods can be applied to model the evolution of the network. Commonly used methods to predict links are similarity-based algorithms[23] (e.g. Adamic-Adar similarity based algorithm[24], resource allocation similarity-based algorithms[25], etc.). Within these algorithms a score is assigned to each pair of nodes. The score is directly defined as the similarity between these nodes[‡‡]. Links are ordered in descending order by their scores. The probability of link existence is high when the score is at the top of the ranking list. There are two groups of similarities: structural similarity and node attribute similarity. Structural similarity takes into account only network structure, while nodes similarity is based only on nodes attributes. Although, most algorithms have been developed for structural similarities because of a lack of node attributes. Using external information as the node attributes, the performance of these algorithms can be enhanced[21]. In addition, as shown by Zheleva et al[26], the combination of network structure, node attributes, and node community features improve link prediction performance. Based on the above, the similarity based algorithms assign scores to all non-observed links utilizing node attributes and rank these scores within communities and between community pairs. In addition, it is important to take into account the possibility of links disappearing from the network. It is assumed that ranked existing links with a lower score, in a given year, are eliminated from the network.

In the case of OD city-pairs, the topology of the network and the attributes including city GDP, population, distance and average airfare between them are known for the base year 2012. Thus, for predicting new links in the OD network it is possible to utilize a weighted similarity based algorithm[27]. Such an approach takes into account not

---

[‡‡] Similarity theorem, introduced by Lin in "An Information-Theoretic Definition of Similarity" (1998), states: the similarity between A and B is measured by the ratio between the amount of information needed to state the commonality of A and B and the information needed to fully describe what A and B are. In other words, two objects are similar to each other if they have many features in common.

only network topology configuration, but the weights of links. These weights could be defined utilizing node attributes. To incorporate communities, set of cities are divided to groups with similar socio-economic indicators.

Thus, for calculating OD air passenger demand network topology, nodes (cities), edges (OD city-pairs), weights (set of socio-economic indicators) and city communities are considered. For each year of the forecast new links are consistently added to the network and old links are eliminated from the network based on the given scenario. Thus, network evolution forecast is a sequential stepwise approach from the base year to the forecast horizon. Based on city attributes, network configuration and communities of the cities, the presented method allows forecasting of air passenger demand network topology according to a socio-economic scenario within the forecast period.

B. Passenger forecast

The forecast of origin-destination air passenger demand between global city pairs consists of a sequential set of discrete "slices" at the time scale up to the forecast horizon. When demand network topology is determined for the forecast period, the number of air passengers on city pairs has to be defined. Based on the assumptions in the study, the passenger number on a given city pair in the base year is determined by the set of socio-economic indicators of these cities and by ATS-specific information (e.g. travel time, aircraft movements). Thus, estimation of the number of air travel passengers on city-pairs has two steps: first, the defining of the number of passengers based on socio-economic indicators and, second, the recalculation of the number of passengers, based on ATS-specific feedback information (Fig. 6).
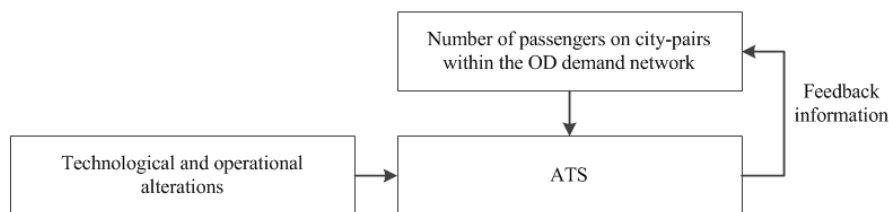

Figure 6: Two step iterative process on estimation number of passengers on a city pairs

In the first step, it is possible to obtain the number of passengers on a given city pair in a given year by searching for the closest set of conditions of a city pair in the base year. In other words, the number of passengers is defined using quantitative analogies (QA) between cross sectional data of a given year and the base year. The mathematical interpretation of calculating the shortest distance between city pairs in a forecast year and city-pairs in the base year is presented as follows:

$$\min_{1 \leq y \leq m} d\,(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \tag{1}$$

Where x_n represents the condition n of city pair x in a given year within the forecast period, y_n represents condition n of city pair y in the base year, and m is the number of connections in the base year. Next, these passenger numbers are passed onto the next model in the chain (routes, aircraft movements and trajectories forecasting). Feedback information from these models (e.g. number of transfers between origin-destination cities) is then taken into account to recalculate the number of passengers on OD city pairs. Using this method, the number of passengers is obtained on every demand connection within a forecast period.

VI. Preliminary results

For preliminary calculations Randers socio-economic scenario[28] from 2012 to 2050, with time slices every 5 years since 2015, is used. Based on cities obtained from the ADI database for the base year 2012, this scenario is decomposed from region level to city level. The obtained datasets contain information of realized undirected demand at city level. The number of cities is constant within the forecast period and contains 4,435 cities from the base year. The conditions taken into account in the study are socio-economic indicators (city GDP, population, oil price, etc.) as well as ATS specific indicators (average air fare, travel time, frequency and number of stops for example). All economic indicators within the study are adjusted to 2005 USD dollars.

The first step is to investigate communities of cities in the set of cities from the base year. Due to this need, clustering of these cities is made to allocate them into groups by similar characteristics. Thus, for the base year, using the normal mixture approach, a clusterization to nine clusters by city populations, city GDP and GDP per

capita has been conducted. Nine clusters cover 'small', 'middle' and 'big' cities by populations and 'poor', 'middle class' and 'rich' cities by GDP. Tab. 1 reflects the number of cities in each cluster and cluster means (cluster centers). For the purposes of the study, cluster names derived from cluster means (of population and per capita GDP) were adopted.

Cluster	Count	Proportion	Cluster mean			Size	Wealth
			Population	GDP, $	GDP p/c, $		
1	1453	0.32191	8,519	3.07E+08	37,134	Very small	Rich
2	1055	0.22774	47,009	3.79E+08	7,728	Small	Poor
3	108	0.02487	824,546	2.71E+10	33,219	Big	Rich
4	417	0.09684	307,440	3.74E+09	12,066	Middle	Middle
5	76	0.01748	5,394,129	7.74E+10	19,767	Megacities	
6	565	0.13312	82,789	2.97E+09	37,009	Small	Rich
7	238	0.05451	1,493,548	1.16E+10	8,032	Big	Poor
8	207	0.04738	278,644	9.73E+09	35,546	Middle	Rich
9	316	0.07615	369,339	1.1E+09	2,743	Middle	Poor

Table 1: Clusters centers and cities distribution among clusters. GDP and GDP per capita indicated here in constant 2005 US dollars.

Over the forecast period, the socio-economic indicators of the cities change. These changes affect the probability of membership of a given city to a certain clusters. This process reveals the changes over time of city distributions within the clusters. It is assumed that during the forecast period, cluster centers remain fixed as in the base year and do not change. Thus, this study introduces 'cluster dynamics'. Cluster dynamics is a method of calculating the probability that a given element (city) will appear within a given cluster at a given point in time. This method is how the cities are allocated to the various clusters in any given forecast year based on socio-economic indicators of cities. Utilizing the Randers scenario, the number of cities in each cluster within the forecast period from 2015 to 2050 have been obtained (see Fig. 5).
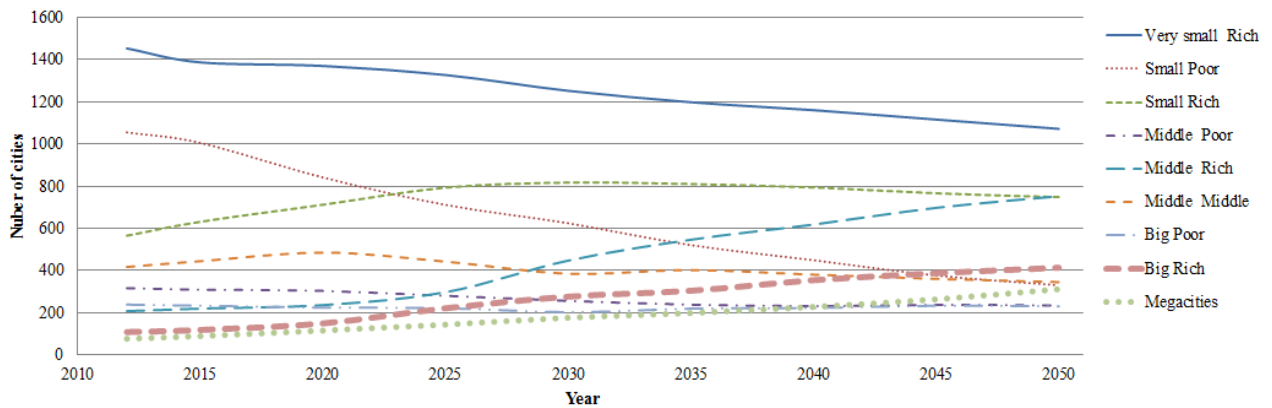

Figure 5: Cluster dynamics based on Randers scenario

For connection predictions within the preliminary calculation, only city attributes have thus far been considered. Based on the obtained data, city attributes are: GDP, population, geographical coordinates. The data for forecasting GDP and population were obtained from Randers scenario. The geographical coordinates are used from the obtained database of the base year (see chapter 3). For the average airfare between cities, a simplified model has been

adopted, as shown by Ghosh[4]. The model utilizes parameters as oil price from a scenario[§§] and great circle distance between cities based on their geographical coordinates.

For connection predictions Newton's gravity model has been adopted to define attractiveness between cities in pairs. Equation (2) presents dependence between two masses reduced by distance between them.

$$F_{ij} = \frac{m_i * m_j}{r_{ij}^2} * G \tag{2}$$

Where F_{ij} represents force between masses i and j; m_i and m_j represent masses of i and j consequently; r_{ij} represents distance between i and j. G is the gravitational constant. In case of city pairs within the preliminary calculation, this gravitational constant assumed to be 1 for every city pair. Based on Eq. (2), attractiveness between cities is presented through the adapted gravity model (see Eq.3).

$$A_{ij} = \frac{GDP_i * Pop_i * GDP_j * Pop_j}{(d_{ij} * AirFare_{ij})^2} \tag{3}$$

Masses of cities i and j are presented as a ratio of multiplication of their GDP and population and are then divided by square of great circle distance between them weighed by average air fare on this connection. This simple model provides a score for every connection within the cluster pairs. For every time slice in the study, scores for all possible connections (9,832,395 connections in one slice) have been obtained. These connections have been allocated in descending order within 45 cluster pairs, based on city memberships within the clusters. As a boundary condition assumptions within this preliminary calculation, a quantile of 0.975 has been adopted. In other words, 2.5% new connections with the highest scores are adding to the network. The quantile of 0.975 has been retrieved from the base year and seems reasonable for preliminary calculations. However, additional research on boundaries must be conducted.

For every cluster pair, quantiles have been obtained from the ordered score lists of each cluster pair in the base year 2012. Thus, for all periods of the forecast, each cluster pair has fixed boundary conditions.
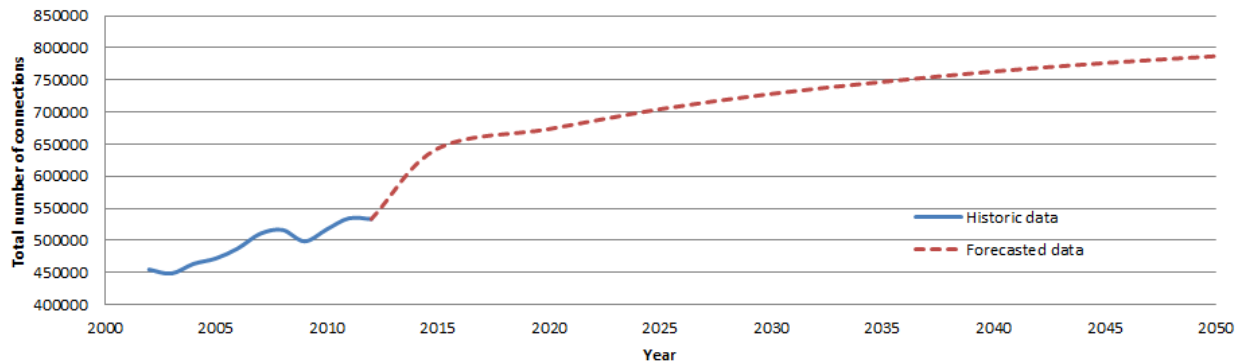

Figure 6: Number of connections in the air passenger demand network based on Randers scenario

Preliminary topology forecasting has been made for each 5-year time slice. The number of connections is shown in Fig. 6. Between 2012 and 2015, a rapid growth in connection numbers can be observed. The reason for this jump can be explained as follows. Firstly, here, the network topology has not yet been considered. Connections were predicted based solely on city attributes and their attractiveness, and the weighed similarity algorithm has not been applied. Secondly, the minimum distance between cities has not yet been set. This causes the appearance of connections between geographically proximate cities. Finally, this preliminary algorithm did not take into account the possibility of eliminating links from the network. The topology forecasting results will be refined in the future taken into account above reasons. However, despite these, the trend of forecasted connections is close to the historical data.

[§§] Randers forecast does not include future oil prices. The Randers forecast in terms of population and GDP is the closest to the "security first" scenario by Hughes and Hillebrand described in their book "Exploring and Shaping International Futures" (2006). This scenario includes future oil prices. Thus, in preliminary calculations oil prices from the "security first" scenario have been used.

For calculating the number of passengers on obtained connections, the quantitative analogies approach has been used. Within preliminary calculations, a reliability analysis of the approach has been made. For this analysis, data from the base 2012 are used. The dataset has been divided into the main and test sets. Among 535,134 city-pairs, 50 connections have been extracted as the test set. As discussed above, a QA approach searches for the minimum distance for a given connection in the forecasted year, and all known connections in the base year. For calculating this distance Eq. (1) has been adapted to:

$$d(x,y) = \sqrt{(GDP_{ix} - GDP_{my})^2 + (GDP_{jx} - GDP_{ny})^2 + (Pop_{ix} - Pop_{my})^2 + (Pop_{jx} - Pop_{ny})^2 + (AF_x - AF_y)^2 + (GCD_x - GCD_y)^2} \qquad (4)$$

Where i and j are cities in the forecasted year in city pair x; m and n are cities in the base year of city pair y; GDP is gross domestic product of a city; Pop is population of a city; AF is average airfare on city pair; GCD is great circle distance between cities in pairs.

The method has been checked versus existing connections, but did not take into account passenger number calculations on new, previously non-existing, connections. Within the preliminary calculations it is assumed that the number of passengers on a city-pair in a forecast year has to be in an interval +/- the number of passengers in the previous year on the same connection. Based on these assumptions, this examination showed R^2 of 0.87 (see Fig. 7). Although imposed boundaries are narrow and require additional study, the method will be pursued for forecasting the number of passengers.
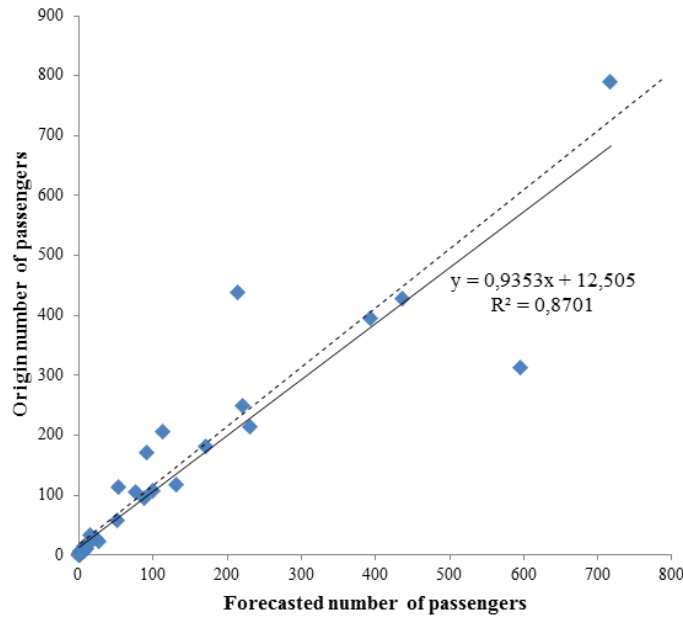

Figure 7: Number of connections in the air passenger demand network based on Randers scenario

VII. Conclusion

The introduced concept of forecasting origin-destination air travel passenger demand is unique among other forecast models. The concept presented above introduces new categorizations that distinguish realized and unrealized air passenger demand. The proposed method to forecast topology changes utilizing a stepwise approach, from the base year to the forecast horizon, combines two groups of link prediction: one based on nodes attributes and one on topology configuration. In addition, communities of nodes have been investigated using clusterization based on socio-economic indicators. For passenger calculations on the predicted connections a quantitative analogies approach has been introduced. Preliminary calculations have yielded promising results.

The method requires further development. Topology forecasting requires additional research and the application of weighted similarity-based algorithms. Currently, few such algorithms exist and it is necessary to determine which one is appropriate for air passenger demand network topology forecasting. In addition, criteria for adding and eliminating connections from/to the network require further study. For passenger forecasting, passenger numbers on

previously non-existing connections must be taken into account. Boundary conditions for passenger number calculations require further investigation.

This concept is a valuable resource for air transportation system research. Within the modular environment *AIRCAST*, the concept aims to provide a valuable foundation for those wishing to perform a wide range of systems analyses e.g. market development, technology assessment and the behavior of air transportation system itself.

Acknowledgments

References

[1] Ghosh, R., Schilling, T., Wicke, K. "Theoretical Framework of Systems Design for the Air Transportation System including an inherently quantitative Philosophy of Scenario Development", *29th Congress of the International Council of the Aeronautical Sciences (ICAS)*, 7.-12, St. Petersburg, Russia, 2014.

[2] Lee, S.D. et al., "Aviation and global change in the 21st century", *Atmospheric Environment*, Vol. 43, 2009, pp. 3520-3537.

[3] Kollmuss, A., Crimmins, M.A., "Carbon Offsetting & Air Travel, Part 2: Non-CO_2 Emissions Calculations", *Stockholm Environment Institute (SEI) discussion paper*, 2009.

[4] Ghosh, R., Terekhov, I., "Future Passenger Air Traffic Modelling: Trend Analysis of the Global Passenger Air Travel Demand Network", *53rd AIAA Aerospace Science Meeting*, Kissimmee, Florida, 2015.

[5] Doucet R., Margaretic, P., Thomas-Agnan, C., Villota, Q., "Spatial dependence in (origin-destination) air passenger flows", *ITEA Annual Conference and Summer School on Transportation Economics (Kuhmo Nectar)*, Toulouse, France, 2014.

[6] Airbus, *Airbus Global Market Forecast 2014-2033*, Blagnac Cedex, France, 2014, http://www.airbus.com/company/market/forecast/?eID=maglisting_push&tx_maglisting_pi1%5BdocID%5D=40733 [cited 19.11.2014].

[7] UK Department for Transport, *UK aviation forecasts 2013*, UK, 2013, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/223839/aviation-forecasts.pdf, [cited 19.11.2014].

[8] Boeing, *Current Market Outlook 2013-2032*, USA, 2013, http://www.boeing.com/assets/pdf/commercial/cmo/pdf/Boeing_Current_Market_Outlook_2013.pdf [cited 19.11.2014].

[9] Japan Aircraft Development Corporation, *Worldwide Market Forecast for Commercial Air Transport 2010-2029,* Chiyoda-ku, Tokyo, Japan, 2014.

[10] L Dray, Evans AD, Reynolds T, Schäfer A, 2010. "Mitigation of Aviation Emissions of Carbon Dioxide: Analysis for Europe," Transportation Research Record, 2177, pp. 17-26.

[11] L Dray, Evans AD, Reynolds R, Schäfer A, Vera-Morales M, Bosbach W, 2014. "Airline Fleet Replacement Funded by a Carbon Tax: an Integrated Assessment," Transport Policy, Special Issue on Aviation and the Environment, 34 (2014), pp. 75-84.

[12] Reynolds, G.T. et al., "Modeling Environmental & Economic Impacts of Aviation: Introducing the Aviation Integrated Modelling Project", *7th AIAA Aviation Technology, Integration and Operations Conference*, Belfast, 2007.

[13] Suryani, E., Chou, S.-Y., Chen C.-H., "Air passenger demand forecasting and passenger terminal capacity expansion: A system dynamics framework", Expert Systems with Applications, Vol. 37, 2010, pp. 2324-2339.

[14] Alam, M.J.B., Karim, D.M., "Air travel demand model for domestic air transportation in Bangladesh", *Journal of Civil Engineering*, Institution of Engineers, Bangladesh, 1998, ISSN 0379-4318, Vol. 26, No.1. pp. 1-13.

[15] Grosche, T., Rothlauf, F., Heinzl, A., "Gravity models for airline passenger volume estimation", *Journal of Air Transport Management*, Vol. 13, 2007, pp. 175-183.

[16] Sabre Airline Solutions, Aviation Data Intelligence, http://www.sabreairlinesolutions.com /home/software_solutions/airports/, [cited 19.11.2014].

[17] UN, National Accounts Main Aggregates Database, http://unstats.un.org/unsd/snaama/dnllist.asp, [cited 19.11.2014].

[18] The World Bank, World Bank Open Data, http://data.worldbank.org/indicator/NY.GDP.MKTP.CD, [cited 19.11.2014].

[19] UN, World population Prospects: The 2012 Revision, http://esa.un.org/unpd/wpp/Excel-Data/population.htm, cited 19.11.2014].

[20] MaxMind, Free World Cities Database, https://www.maxmind.com/en/worldcities, [cited 19.11.2014].

[21] Our Airports, http://ourairports.com/data/, [cited 19.11.2014].

[22] OpenFlights, Airport database, http://openflights.org/data.html, [cited 19.11.2014].

[23] Lü, L., Zhou, T., "Link prediction in complex networks: A survey", *Physica A*, Vol. 390, 2011, pp. 1150-1170.

[24] Adamic, L.A., Adar, A., "Friends and neighbors on the web", *Social Networks*, Vol. 23, Issue 3, July 2003, pp. 211-230

[25] Zhou, T., Lü, L., Zhang, Y.-C., "Predicting missing links via local information", *The European Physical Journal B*, Vol. 71, Issue 4, October 2009, pp. 623-630.

[26] Zheleva, E., Golbeck, J., Kuter, U., "Using Friendship Ties and Family Circles for Link Prediction", *Advances in Social Network Mining and Analysis Lecture Notes in Computer Science*, Vol. 5498, 2012, pp. 97-113.

[27] Lü, L., Zhou, T., "Link prediction in weighted networks: The role of weak ties", *A Letters Journal Exploring the Frontiers of Physics*, Vol. 89, 2012, 18001.

[28] Randers, J., "2052: A global Market Forecast for the Next Forty Years", *Chelsea Green Publishing*, White River Junction, Vermont, 2012.

American Institute of Aeronautics and Astronautics