Structural Equation Modelling Based Data Fusion for Technology Forecasting: A National Research and Education Network Example

Leon Staphorst[1], Leon Pretorius[2], Tinus W. Pretorius[2]
[1]Council for scientific and Industrial Research, Brummeria, Pretoria, South Africa
[2]Department of Engineering and Technology Management, University of Pretoria, Pretoria, South Africa

Abstract--**This paper presents an example model instantiation of Staphorst, Pretorius and Pretorius' framework for Structural Equation Modelling (SEM) based Data Fusion (DF) for Technology Forecasting (TF) in the National Research and Education Network (NREN) technology domain. The paper's example NREN model instantiation is constructed through deductive reasoning from knowledge gained during action research in the South African National Research Network (SANReN), as well as secondary data from TERENA's NREN compendiums for global NREN infrastructure and services trends. A variety of technology related measurements are employed in the example NREN model instantiation as indicators for technology related model constructs, such as the level of core network traffic in an NREN. Indicators for context related model constructs include, amongst others, the range of institutions an NREN is mandated to connect. For confirmatory purposes the secondary data published by TERENA in its yearly NREN compendium series is then used in the Partial Least Squares (PLS) regression analysis to determine the indicator loadings and path coefficients of the example NREN model instantiation. A reliability and validity analysis of the example NREN model instantiation is also considered.**

I. INTRODUCTION AND RESEARCH METHOD

The survival, growth and profitability of firms that play in technology based products and services markets depend highly on their ability to monitor current, as well as predict future technological changes [10]. Building a solid and sustainable technological base that can withstand or adapt to rapidly changing market requirements necessitate these firms to effectively and efficiently manage technological changes, both internally and externally [10]. Core to this technology management challenge is the Technology Intelligence (TI) process, which spans the capturing technology related data, converting this data into information (by determining relational connections) and then refining this information to produce knowledge that can guide strategic decision makers [7][4]. Measureable sources of technology related data that allow for the direct characterisation and evaluation of technologies throughout their life cycle are defined as technology indicators [4]. Technology Forecasting (TF) entails the creation of knowledge for strategic decision-making in areas such as investing or divesting in certain technologies by means of a forward-looking scrutinization of the information distilled from a set of technology indicators [9].

Staphorst, Pretorius and Pretorius [14] postulated that the TI process, which involves the capturing technology related data, conversion of this data into information by determining relational connections and refining information to produce knowledge that can guide strategic decision making, can be viewed as an instance of context sensitive Data Fusion (DF). DF was developed in the military domain for the generation of quality tactical knowledge through the multi-layered processing of sensor data [18]. It is a framework capable of performing multi-layered refinement of estimates of problem variables from multiple measurements, either directly or indirectly observable [16]. Context in the domain of DF is defined as a set of relational connections [16] which can be used in each level of the DF process in order to refine data alignment and association, as well as during situation state estimation [16]. Staphorst, Pretorius and Pretorius [14] argued that the application of context sensitive DF during the TI process could increase the quality of the generated technology related knowledge by increasing the accuracy of the relational connections defined between technology indicators. Staphorst, Pretorius and Pretorius [14] went on to propose a framework for a Structural Equation Modelling (SEM) based implementation of context sensitive DF of technology indicator data for the purpose of performing TF. SEM is a second generation statistical regression technique that allows for the simultaneous modelling of relationships among multiple dependent and independent constructs, which can be latent or observable in nature. Construction of Staphorst, Pretorius and Pretorius' framework, which is described in detail in [15], entailed the merger of Steinberg's [16][17] implementation of context sensitive DF using SEM, with Sohn and Moon's [12] use of SEM to implement TF. Unlike most TF approaches [12], the framework proposed by Staphorst, Pretorius and Pretorius [14] not only caters for complex and hierarchical structural relationships between technology indicators and TF output metrics, but also allows for non-linear and non-Gaussian factors and cyclical dependencies amongst model variables, which can be either latent or directly observable.

An NREN is a specialised broadband network connectivity and service provider that explicitly caters for the needs of the research and education communities of a country [2]. In some instances, NRENs also service the needs of other public sector entities, such as hospitals, municipalities and libraries. Typically, one NREN is present per country (for example SANReN [3] in South Africa and the Joint Academic Network (JANET) in the United Kingdom), although separate NREN entities could potentially exist to service distinct in-country research and education sectors (for example the Energy Sciences Network (ESnet) and Internet2 in the United States) [2]. NREN's are built primarily on fibre optic cabling infrastructure and provide researchers, educators and students with unparalleled connectivity speeds and advanced services at a fraction of the price of commercial network providers [2]. These networks are currently experiencing rapid technology driven changes, resulting in evolving business models, innovative infrastructure solutions and service offerings, as well as increased international collaboration [1][2].

This paper commences with an overview of an evolved version of the original framework [14] proposed by Staphorst, Pretorius and Pretorius, followed by the application of the framework to the technology domain of National Research and Education Networks (NRENs). The evolved framework, as well as the example model instantiation in the NREN technology domain, was first proposed in [15]. The proposed NREN example model instantiation was constructed through deductive reasoning using insights gained through action research in the South African National Research Network (SANReN) [3], as well as from TERENA's NREN compendium for 2012 on global NREN infrastructure and services trends [2]. A number of research propositions related to this example model instantiation are defined. Next, a quantitative evaluation of the example model instantiation is performed using longitudinal data extracted from TERENA's NREN compendiums for 2011 [1] and 2012 [2]. This includes an evaluation of the research propositions defined. Lastly, the paper presents an evaluation of the reliability and the validity of the example NREN model instantiation.

II. FRAMEWORK FOR SEM BASED DF FOR TF

By noting that SEM is capable of the simultaneous modelling of relationships among multiple dependent and independent constructs, Steinberg [16][17] postulated that SEM is one potential statistical tool that lends itself naturally to implement DF. Moreover, based on the following argumentation Steinberg [12] showed that SEM allows for the inclusion of context sensitivity during the solving of DF inferencing problems: Firstly, Steinberg [16][17] defined a situation, or a context, as a set of relationships, where a relationship can be viewed as a specific instantiated relation. In general, context is used in DF inferencing problems in order to refine ambiguous estimates, explain available data and constraint processing during data acquisition, cueing or fusion [16][17]. Next, Steinberg harmonized DF and SEM terminology by noting that DF problem variables are in fact SEM endogenous constructs, context variables can be viewed as SEM exogenous constructs and classic DF sensor measurements are the reflective and formative indicators present in SEM [16][17].

Nyberg and Plamgren [8] describe technological indicators as indices or statistical data that allow for the direct characterisation of technology throughout their life cycles in order to allow decision makers to take strategic actions. According to Grupp [5], such indicators can in general be divided into the following three major categories based on their intended function: input indicators, byput indicators and output indicators [8][5]. Grupp [5] states that input indicators are variables related to drivers of technological progress, byput indicators are variables that are related to sub-phenomena of the technological progress and output indicators are variables related to the qualitative, quantitative or value-rated progress in process or product development [8].

Sohn and Moon showed in [12] that SEM can be used as an effective regression technique to evaluate a multi-layered hierarchal model through progressive aggregations and refinements of input technology indicator data in order to produce a reliable statistical estimate of the Technology Commercialization Success Index (TCSI) TF output metric [15]. By extending Soon and Moon's [12] use of SEM for TF and Steinberg's use of SEM to implement context sensitive DF [16], Staphorst, Pretorius and Pretorius [14][15] developed, using inductive reasoning, the framework depicted in Fig 1. In this SEM based DF for TF framework multi-layered aggregation and refinement of technology and context related information is accomplished by the processing performed at DF Levels 0 through N-1, where N is user selected. The number of levels N will be determined not only by the complexity of the technology domain under consideration, but also by time and cost constraints of collecting the technology indicator data [15]. Furthermore, potential diminishing returns resulting from additional levels of aggregation and refinement will also be determining factors in defining N [15].

In this framework input technology indicators [8][5] and context related indicators [16] are used as inputs to technology related endogenous constructs and context related exogenous constructs, respectively. Note that the use of bi-directional interconnections between indicators and constructs, as well as between multiple constructs, is based on the SEM path diagram conventions defined in [13]. This illustrates that positive or negative correlation can exist between constructs, as well as the fact that indicators can be either reflective or formative in nature.

To gain insight into the functioning of this framework, consider the aggregation and refinement that occur in progressing from DF Level 0 to DF Level 1: Regression analysis outputs generated for the technology related exogenous constructs at DF Level 0 contribute formatively of reflectively to technology related endogenous constructs at DF Level 1. Regression analysis outputs for the context related exogenous constructs of DF Level 0 contribute to context related exogenous and technology related endogenous constructs at DF Level 1. The regression analysis results produced at DF Level 1 for context related exogenous constructs can also contribute to technology related endogenous constructs at this same level. Technology indicators for the technology related constructs at DF Level 1 could potentially be selected as the TF output metrics, or could simply be byput technology [8][5] indicators if additional DF levels are required for further aggregation and refinement. The aggregation and refinement achieved by moving from DF Level x-1 to DF Level x, for x = 1, 2, 3,…, N-1, follows a similar interconnection structure as the progression from DF Level 0 to DF Level 1, with the exception that now constructs at DF Level x-1 contribute to constructs at DF Level x.

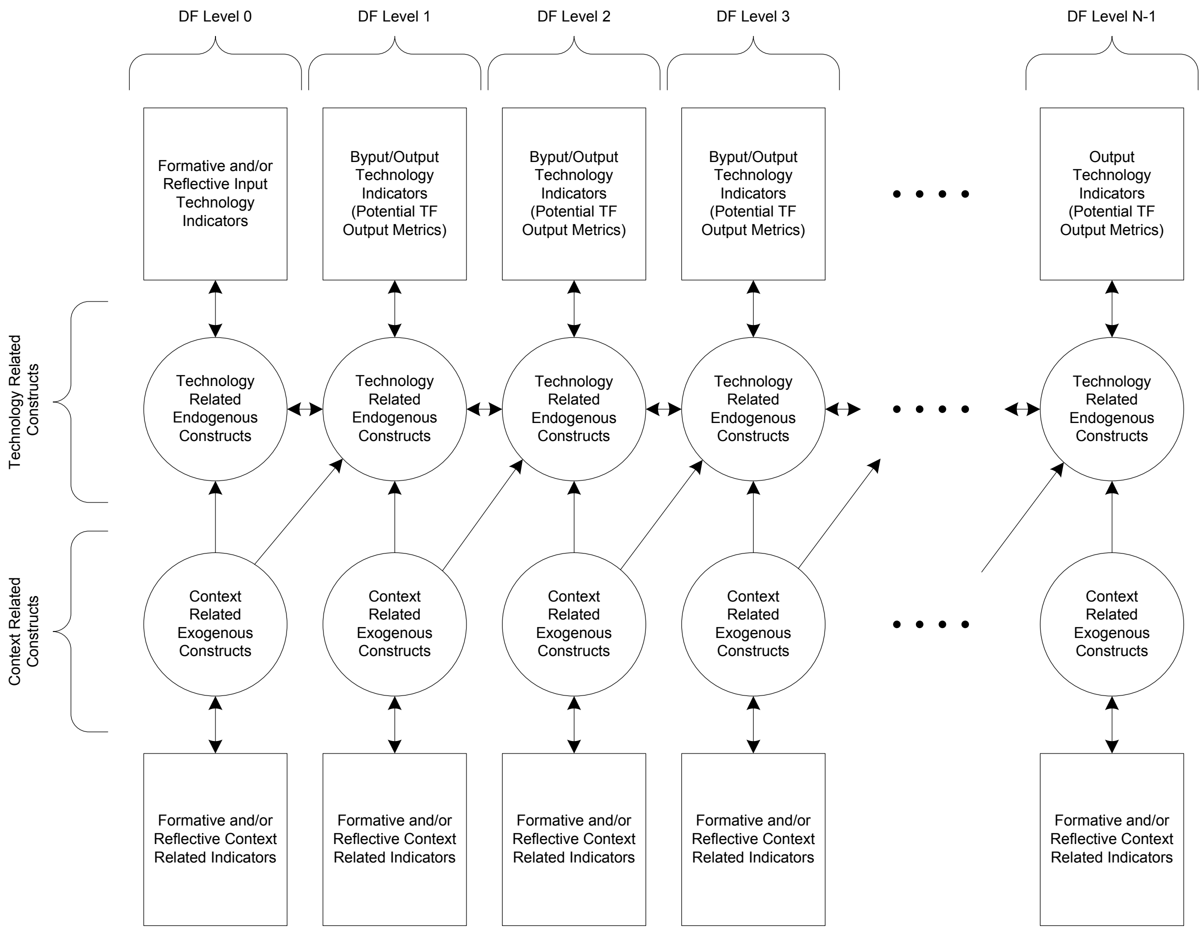

Fig 1: Proposed Framework for SEM Based DF for TF [14][15]

III. EXAMPLE MODEL INSTANTIATION FOR THE NREN TECHNOLOGY DOMAIN

NREN's are frequently used as incubators for the development of new networking technologies and services [2]. Hence, NREN's contribute significantly to the creation of new Internet based business ventures, innovative business models and game changers in the way society works and plays. For example, Facebook and Google have their roots within the NREN environments of Harvard University and Stanford University, respectively.

TERENA is an international community of practice for the global community of NREN's. It offers a platform for NREN's to collaborate and openly share knowledge on networking technologies, services and infrastructure. TERENA performs an extensive yearly survey amongst the global NREN community in order to determine current technology and services trends. The results and interpretation of these surveys are then openly published as part of TERENA's NREN compendium series. The NREN model instantiation example detailed in the following subsection was created using insights captured in TERENA's NREN compendium for 2012 [2], as well as knowledge gained through action research [6] performed by the authors during their involvement with the management and operations of SANReN [3]

A. NREN Model Instantiation Overview

Fig. 2 presents the proposed example NREN technology domain model instantiation of the framework discussed in Section II. This example model instantiation employs N=3 DF levels. Level 0, Level 1 and Level 2 focus on NREN infrastructure, NREN services and NREN reach, respectively.

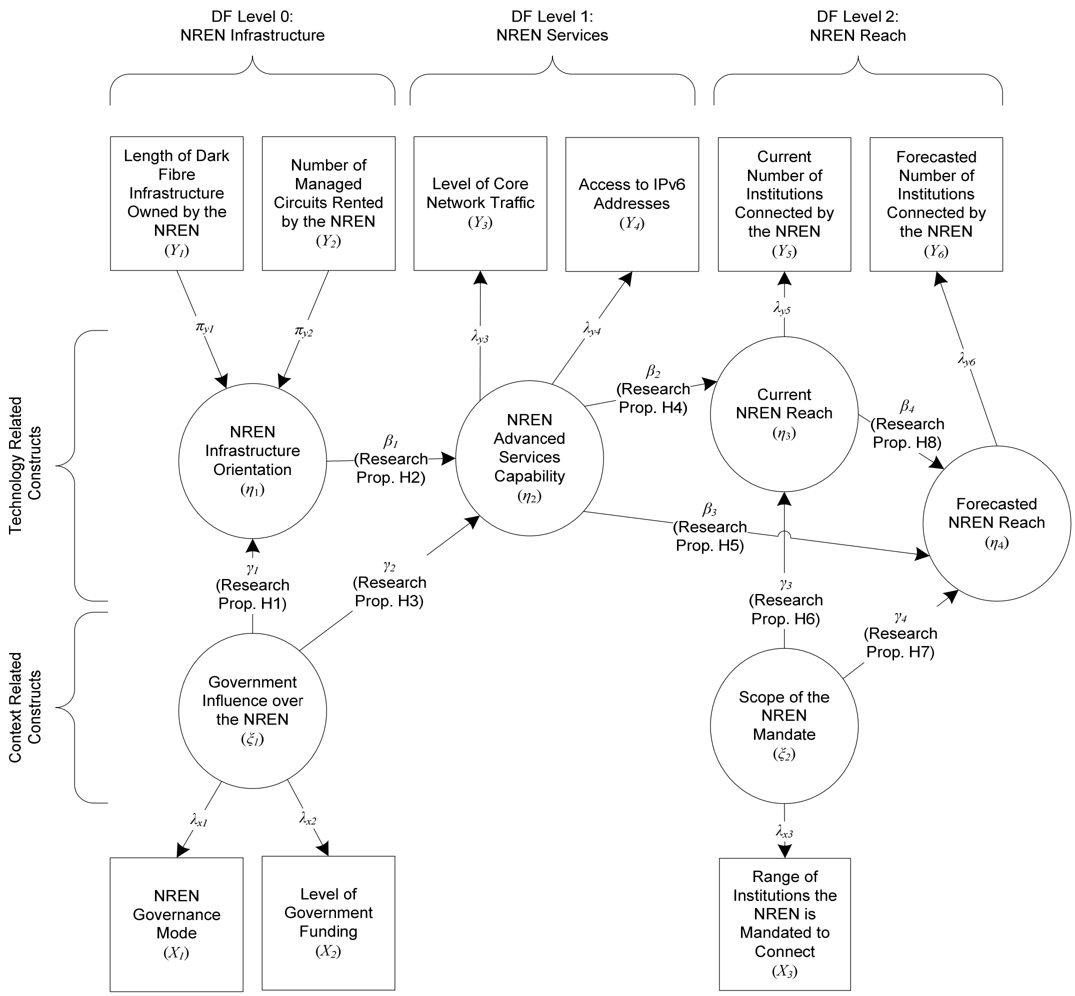

Fig 2: Model Instantiation Example for the NREN Technology Domain [15]

At Level 0 of the example model instantiation a single technology related endogenous construct, namely *NREN Infrastructure Orientation* (η_1), is defined. The purpose of this construct is to model the extent to which the NREN prefers investing in ownership over leasing of fibre optical cable infrastructure [2]. It is postulated that this construct will be related to two formative input technology indicators (i.e. both indicators jointly represent the construct) that measure the length of dark fibre [2] owned optical cabling infrastructure (denoted as Y_1 with indicator loading π_{y1}) and number of rented managed circuits (denoted as Y_2 with indicator loading π_{y2}), respectively. Also at Level 0 a single context related exogenous construct entitled *Government Influence over the NREN* (ξ_1) is defined, with two reflective indicators (i.e. each indicator is capable of individually representing the construct) that measure the NREN governance mode (denoted as X_1 with indicator loading λ_{x1}) and level of government funding to the NREN (denoted as X_2 with indicator loading λ_{x2}), respectively. NREN governance mode can range from full government driven governance through to no government driven governance [2][13]. The postulated positive relation between *Government Influence over the NREN* (ξ_1) and *NREN Infrastructure Orientation* (η_1) is reflected by the path coefficient γ_1.

Level 1 of the example model instantiation defines a single endogenous technology related construct entitled *NREN Advanced Services Capability* (η_2), which embodies the NREN's capability to provide a suite of advanced NREN services (i.e. services other than commodity Internet services, e.g. Lightpaths and Science Gateways [2]). While no exogenous context related construct is defined for this level, it is postulated that the Level 0's constructs that capture the level of government influence over the NREN, as well as the NREN's preference to invest in owned over leased fibre optic cable infrastructure, will be positively related to the NREN's ability to deliver advanced services [2]. These relationships are captured in the SEM model of Fig 2 by means of path coefficients γ_2 and β_1, respectively. This postulation is based on the reasoning that less government control and more freedom over the usage of its fibre optic cable infrastructure will enhance an NREN's capability to develop and deliver advanced services.

The *NREN Advanced Services Capability* construct of Level 1 is represented by two reflective byput technology indicators, namely the level of network traffic measured in the core infrastructure of the NREN (denoted as Y_3 with indicator loading λ_{y3}) and the percentage of users with access to Internet Protocol version 6 (IPv6) addresses (denoted as Y_4 with indicator loading λ_{y4}) [2]. The reasoning for using the latter technology indicator is that a shortage of classic IPv4 addresses has been identified as a barrier in the development of new services [1].

Level 2 in the example model instantiation focuses on the reach of the NREN network, which is frequently used as a proxy to measure the impact that an NREN creates in its beneficiary communities [2][3]. A single context related exogenous construct entitled *Scope of the NREN Mandate (ξ_2)* is defined. This construct is represented by a single reflective measurement indicator that captures the number of distinct types of institutions the NREN is allowed to provide connectivity to (denoted as X_3 with indicator loading λ_{x3}), ranging from a very narrow NREN mandate that only allows higher education and research institutions to be connected, through to a very broad NREN mandate that allows all public (and even some private sector research entities) to be connected to the NREN [2]. In terms of technology related endogenous constructs the *Current NREN Reach (η_3)* and *Forecasted NREN Reach (η_4)* constructs are proposed for Level 2, with the former represented by a single byput technology indicator that counts the number of institutions currently connected to the NREN [2] (denoted as Y_5 with indicator loading λ_{y5}). The *Forecasted NREN Reach* construct's single reflective output technology indicator, which measures the forecasted number of institutions connected to the NREN (denoted as Y_6 with indicator loading λ_{y6}), is used as the example model's TF output metric.

It is postulated that Level 1's *NREN Advanced Services Capability* construct will be positively related to both the current and forecasted NREN reach constructs (indicated through path coefficients β_2 and β_3, respectively), as it has been observed in SANReN [3] that the demand for NREN connectivity from unconnected entities grows as an NREN's portfolio of advanced services expands, since this is perceived by potential beneficiary institutions as an indicator of the maturity and stability of the NREN [3]. It is also postulated that *Scope of the NREN Mandate (ξ_2)* is positively related to the current and forecasted NREN reach constructs (indicated through path coefficients γ_3 and γ_4, respectively). Lastly, in the case of SANReN it has been observed that the current NREN reach is positively related to the forecasted NREN reach [3] (indicated by path coefficient β_4), as one would innately expect.

B. Research Propositions Emanating from the Example NREN Model Instantiation

The postulated relationships between constructs in Fig.2's example NREN model instantiation give rise to the set of research propositions below which are evaluated in Section IV.C. These research propositions' association with the various paths defined in the example NREN model instantiation is detailed in Fig. 2, as well as Table 3.

- **Research Proposition H1:** The NREN infrastructure orientation is positively related to the level of government influence over the NREN.
- **Research Proposition H2:** The advanced services capability of the NREN is positively related to the NREN's preference to invest in owned fibre optic cable infrastructure over leased fibre optic cable infrastructure.
- **Research Proposition H3:** The advanced services capability of the NREN is positively related to the level of government influence over the NREN.
- **Research Proposition H4:** The current NREN reach is positively related to the advanced services capability of the NREN.
- **Research Proposition H5:** The forecasted NREN reach is positively related to the advanced services capability of the NREN.
- **Research Proposition H6:** The current NREN reach is positively related to the scope of the NREN's mandate.
- **Research Proposition H7:** The forecasted NREN reach is positively related to the scope of the NREN's mandate.
- **Research Proposition H8:** The forecasted NREN reach is positively related to the current NREN reach.

IV. SEM REGRESSION RESULTS FOR THE EXAMPLE NREN MODEL INSTANTIATION

Secondary data from TERENA's NREN compendiums for 2011 [1] and 2012 [2] were used to determine Fig. 2's indicator loadings and path coefficients through PLS regression analysis. Table 1 below summarises the composition of the NREN model instantiation indicator data using the secondary data extracted from these TERENA NREN compendiums [2][1]. A total of 59 NRENs responded to TERENA's survey to collect data for the 2011 compendium [1], while 54 NRENs responded to the 2012 survey [2]. The original 2011 and 2012 surveys distributed by TERENA to NRENs are available from [1] and [2], respectively.

TABLE 1: TECHNOLOGY AND CONTEXT RELATED INDICATOR DATA COMPOSITION

Technology or Context Related Indicator	***Indicator Composition***
NREN Governance Mode (X_1)	Extracted from the online profiles of the respondent NRENs of the 2011 compendium [1] using following the scaling: The NREN is a government agency or part of a ministry = 3 Government appoints at least half of the NREN's governing body = 2 Indirect relationship between the NREN and government = 1 No formal relationship between the NREN and government = 0
Level of Government Funding (X_2)	Level of government funding (as a percentage of total funding) received by respondent NRENS, as summarised in Graphs 6.4.2 and 6.4.3 in the 2011 NREN compendium [1]
Range of Institutions the NREN is Mandated to Connect (X_3)	Sum of the institution types in Table 2.2.1 of the 2011 NREN compendium [1] supported by respondent NRENs
Length of Dark Fibre Infrastructure Owned by the NREN (Y_1)	Total length of dark fibre [in kilometres] owned by respondent NRENs as summarised in Table 3.6.3 of the 2011 NREN compendium [1]
Number of Managed Circuits Rented by the NREN (Y_2)	Total number of managed circuits rented by respondent NRENs as summarised in Table 3.3.2 of the 2011 NREN compendium [1]
Level of Core Network Traffic (Y_3)	Annual level (measured in terabytes per year) of traffic sent on to the backbone networks of respondent NRENs, as measured by T_1+T_4 in Graphs 4.2.1 and 4.2.2 in the 2011 NREN compendium [1]
Access to IPv6 Addresses (Y_4)	Sum of the percentage of users from respondent NRENs that have both IPv4 and IPv6 addresses with the percentage of users that only have IPv6 addresses, as listed in Tables 4.6.1 and 4.6.2 in the 2011 NREN compendium [1]
Current Number of Institutions Connected by the NREN (Y_5)	Sum of the institutions in Table 2.2.1 of the 2011 NREN compendium [1] connected by respondent NRENs
Forecasted Number of Institutions Connected by the NREN (Y_6)	Sum of the institutions in Table 2.2.1 of the 2012 NREN compendium [2] connected by respondent NRENs

In this study the SmartPLS [11] freeware software package was employed to realise the example NREN model instantiation of Fig. 2 and calculate all loadings and path coefficients through PLS regression. SmartPLS was configured to normalise all indicator data, as a variety of scaling approaches and ranges was used by TERENA in collecting the original data. SmartPLS was also used to evaluate the reliability and validity test criteria defined in [14][15] with the results discussed in Section IV.B. Note that only 27 NRENs provided all of the survey inputs in order to calculate the indicator inputs according to Table 1. Hence missing data was flagged and SmartPLS configured to use a mean replacement algorithm to compensate for this [11].

A. Measurement Indicator and Path Coefficient Results

The reporting of the PLS regression results for the example NREN model, presented in the following subsections, was based on the reporting standard defined by Vinzi, Chin, Henseler and Wang [19]. According to this reporting standard, the PLS regression results for the measurement portion of the SEM path diagram, consisting of the loadings for all of the measurement indicators in the model, are reported first, followed by the PLS regression results for the structural portion of the SEM path diagram, consisting of the path coefficients for all interrelationships between constructs.

1. Measurement Portion SEM Regression Results

The indicator loadings for the measurement portion of the example NREN model instantiation, determined using SmartPLS [11], are listed in Table 2. Although these loadings were not used directly in order to evaluate the research propositions stated in Section III.B, a detailed investigation thereof was crucial in order to determine those reflective indicators that did not comply with the minimum Indicator Reliability level of 0.4 (see Section IV.B.1). The results given in Table 2 constitute the final indicator loadings determined following the removal of the *Access to IPv6 Addresses* unreliable reflective indicator, which was revealed during a first-run PLS regression SEM analysis. Removal of this unreliable reflective indicator resulted in improved Construct Reliability for their associated latent constructs (see Section IV.B.1).

TABLE 2: MEASUREMENT PORTION INDICATOR LOADING RESULTS

Endogenous and Exogenous Constructs	***Construct Type***	***Indicator Type***	***Measurement Indicators***	***Loadings***
Government Influence of the NREN (ξ_1)	Latent	Reflective	NREN Governance Mode (X_1)	λ_{x1} = 0.7402
		Reflective	Level of Government Funding (X_2)	λ_{x2} = 0.7487
Scope of the NREN Mandate (ξ_2)	Observable	Reflective	Range of Institutions the NREN is Mandated to Connect (X_3)	λ_{x3} = 1.00
NREN Infrastructure Orientation (η_1)	Latent	Formative	Length of Dark Fibre Infrastructure Owned by the NREN (Y_1)	π_{y1} = 0.3888
		Formative	Number of Managed Circuits Rented by the NREN (Y_2)	π_{y2} = 0.9111
NREN Advanced Services Capability (η_2)	Latent	Reflective	Level of Core Network Traffic (Y_3)	λ_{y3} = 1.0
		Reflective	Access to IPv6 Addresses (Y_4)	Excluded as loading is less than 0.4
Current NREN Reach (η_3)	Observable	Reflective	Current Number of Institutions Connected by the NREN (Y_5)	λ_{y5} = 1.0
Forecasted NREN Reach (η_4)	Observable	Reflective	Forecasted Number of Institutions Connected by the NREN (Y_6)	λ_{y6} = 1.0

2. Structural Portion SEM Regression Results

The path coefficients for the structural portion of the example NREN model instantiation, which were determined using SmartPLS [11], are listed in Table 3. Significance testing for these path coefficients, based on asymptotic t-statistics, is presented in Section IV.B.2. These path coefficients and their associated significance test results were used in Section IV.C to evaluate the research propositions listed in Section III.B.

B. Reliability and Validity of the Example NREN Model Instantiation

Similar to the reporting standard for SEM indicator loading and path coefficient results Vinzi, Chin, Henseler and Wang [19] suggest that the reporting of reliability and validity test results first considers the measurement portion, which include Indicator Reliability, Construct Reliability and Convergent Validity [14][15]. This is then followed by the structural portion, which include Coefficients of Determination, Path Coefficient Significance and Predictive Validity [14][15].

1. Measurement Portion Reliability and Validity Test Results

This subsection details the reliability and validity test results for the measurement portion of the SEM for the example NREN model instantiation, based on the metrics defined in [15] and determined using SmartPLS [11]. Table 4 presents the Indicator Reliability judgement, Construct Reliability and Convergent Validity test results.

The Indicator Reliability test results revealed that the *Access to IPv6 Addresses (Y_4)* reflective indicator exhibited a loadings less than 0.4 during a first-run PLS regression SEM analysis. As a result, this unreliable reflective indicator was removed from all subsequent SEM analyses. All formative indicators were retained, regardless of their loadings [14][15].

Construct Reliability tests considered both the classic Cronbach's Alpha metric and the more contemporary Composite Reliability measure [14][15]. This study's final judgment on the adequacy of a set of reflective indicators to measure their related latent construct was based on the requirement that the Composite Reliability measure needs to exceed a minimum level of 0.6 [19]. As is clear from Table 4 the only latent construct with reflective indicators present in the model after pruning unreliable indicators was *Government Influence of the NREN (ξ_1)*, which complied with this requirement for Composite Reliability.

TABLE 3: STRUCTURAL PORTION PATH COEFFICIENT RESULTS

SEM Path for the Example NREN Model Instantiation	***Research Proposition***	***Path Coefficient***
Government Influence of the NREN (ξ_1) → NREN Infrastructure Orientation (η_1)	H1	$\gamma_1 = 0.1659$
NREN Infrastructure Orientation (η_1) → NREN Advanced Services Capability (η_2)	H2	$\beta_1 = 0.5097$
Government Influence of the NREN (ξ_1) → NREN Advanced Services Capability (η_2)	H3	$\gamma_2 = -0.2073$
NREN Advanced Services Capability (η_2) → Current NREN Reach (η_3)	H4	$\beta_2 = 0.0704$
NREN Advanced Services Capability (η_2) → Forecasted NREN Reach (η_4)	H5	$\beta_3 = -0.0095$
Scope of the NREN Mandate (ξ_2) → Current NREN Reach (η_3)	H6	$\gamma_3 = 0.1586$
Scope of the NREN Mandate (ξ_2) → Forecasted NREN Reach (η_4)	H7	$\gamma_4 = -0.0229$
Current NREN Reach (η_3) → Forecasted NREN Reach (η_4)	H8	$\beta_4 = 0.9964$

TABLE 4: INDICATOR RELIABILITY, CONSTRUCT RELIABILITY AND CONVERGENT VALIDITY TEST RESULTS

Endogenous and Exogenous Constructs	***Measurement Indicators***	***Indicator Reliability Judgement***	***Construct Reliability***		***Convergent Validity***
			Cronbach's Alpha	***Composite Reliability***	
Government Influence of the NREN (ξ_1)	NREN Governance Mode (X_1)	Included	0.1958	$\rho_{\xi,1} = 0.7132$	$AVE_{\xi,1} = 0.5542$
	Level of Government Funding (X_2)	Included			
Scope of the NREN Mandate (ξ_2)	Range of Institutions the NREN is Mandated to Connect (X_3)	Included	Tests not applicable: This construct is directly observable [15]		
NREN Infrastructure Orientation (η_1)	Length of Dark Fibre Infrastructure Owned by the NREN (Y_1)	Included	Tests not applicable: This construct is directly observable and has formative indicators [15]		
	Number of Managed Circuits Rented by the NREN (Y_2)	Included			
NREN Advanced Services Capability (η_2)	Level of Core Network Traffic (Y_3)	Included	Tests not applicable: While this construct was defined to be latent, exclusion of Y_4 resulted in it being treated as directly observable [15]		
	Access to IPv6 Addresses (Y_4)	Excluded			
Current NREN Reach (η_3)	Current Number of Institutions Connected by the NREN (Y_5)	Included	Tests not applicable: This construct is directly observable [15]		
Forecasted NREN Reach (η_4)	Forecasted Number of Institutions Connected by the NREN (Y_6)	Included	Tests not applicable: This construct is directly observable [15]		

Convergent Reliability, which was determined through the AVE metric [14][15], measured the variance of each latent construct's reflective indicators (as captured by the construct itself) relative to the total measured variance. Measured against the study's elected threshold value of 0.5 for this metric, it can be concluded from Table 4's results that the reflective indicators of the only remaining latent construct *Government Influence of the NREN* (ξ_1) exhibited a sufficient AVE level, indicating that for this construct the majority of the total variance measured was due to indicator variance and not due to measurement error.

An evaluation of the Discriminant Validity for the example NREN model instantiation was redundant for the following reason: The requirement for Discriminant Validity is that the square root of each latent construct's AVE exceeds its correlation with all other latent constructs [14][15]. Since the example NREN model instantiation only had one remaining latent construct with reflective indicators, namely *Government Influence of the NREN* (ξ_1), after the pruning of the unreliable indicator *Access to IPv6 Addresses* (Y_4), this test was unnecessary.

2. Structural Portion Reliability and Validity Test Results

The results for the reliability and validity tests for the structural portion of example NREN model instantiation, based on the metrics defined in [15], are presented in this subsection. Table 5 details the Path Coefficient test results, while Table 6 considers the Coefficients of Determination and Predictive Validity test results, all obtained using SmartPLS [11].

From Table 5's Path Coefficient Significance test results, obtained using SmartPLS's bootstrapping function [11] configured for a resampling size of 1000, it is clear that the following paths exhibited *p*-values (calculated using the $t_{(999)}$ asymptotic t-statistic distribution) larger than the maximum acceptable significance level of $\alpha = 0.10$ and were therefore deemed insignificant [15]:

- *Government Influence of the NREN* (ξ_1) → *NREN Infrastructure Orientation* (η_1)
- *Government Influence of the NREN* (ξ_1) → *NREN Advanced Services Capability* (η_2)
- *NREN Advanced Services Capability* (η_2) → *Current NREN Reach* (η_3)
- *NREN Advanced Services Capability* (η_2) → *Forecasted NREN Reach* (η_4)

The Coefficients of Determination test results given in Table 6 revealed that the interrelationships between the *NREN Infrastructure Orientation* (η_1) and *Current NREN Reach* (η_3) endogenous latent constructs and other related constructs did not produce explained variances exceeding the minimum level of 10% [15]. Also, interrelationships with the *Forecasted NREN Reach* (η_4) endogenous latent construct were deemed to be strong, since the R^2 for this construct exceeded 0.7 [15]. Interrelationships with the NREN Advanced Services Capability (η_2) endogenous latent construct were viewed as weak; since the R^2 for this construct was lower than 0.3 [15].

TABLE 5: PATH COEFFICIENT SIGNIFICANCE TEST RESULTS

SEM Path for the Example NREN Model Instantiation	*Asymptotic t-Statistic*	*Calculated p-Value*	*Significance Judgement*		
			$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
Government Influence of the NREN (ξ_1) → NREN Infrastructure Orientation (η_1)	1.2926	0.196	No	No	No
NREN Infrastructure Orientation (η_1) → NREN Advanced Services Capability (η_2)	3.3564	0.001	Yes	Yes	Yes
Government Influence of the NREN (ξ_1) → NREN Advanced Services Capability (η_2)	0.8783	0.380	No	No	No
NREN Advanced Services Capability (η_2) → Current NREN Reach (η_3)	1.5468	0.122	No	No	No
NREN Advanced Services Capability (η_2) → Forecasted NREN Reach (η_4)	1.3593	0.174	No	No	No
Scope of the NREN Mandate (ξ_2) → Current NREN Reach (η_3)	2.3155	0.021	No	Yes	Yes
Scope of the NREN Mandate (ξ_2) → Forecasted NREN Reach (η_4)	2.3489	0.019	No	Yes	Yes
Current NREN Reach (η_3) → Forecasted NREN Reach (η_4)	39.0519	0.000	Yes	Yes	Yes

TABLE 6: COEFFICIENTS OF DETERMINATION AND PREDICTIVE VALIDITY TEST RESUTLS

Technology or Context Related Indicator	*Coefficients of Determination (R^2)*	*Predictive Validity (Q^2)*	
		Cross-validated Communality (H^2)	*Cross-validated Redundancy (F^2)*
Government Influence of the NREN (ξ_1)	Test not applicable: Exogenous variable	0.5542	0
Scope of the NREN Mandate (ξ_2)	Test not applicable: Exogenous variable	1.0	0
NREN Infrastructure Orientation (η_1)	0.0275	0.4906	0.0137
NREN Advanced Services Capability (η_2)	0.2677	1.0	0.0079
Current NREN Reach (η_3)	0.0287	1.0	0.0035
Forecasted NREN Reach (η_4)	0.9852	1.0	-0.0011

A review of the Predicative Validity test results for the *Forecasted NREN Reach (η_4)* construct, directly observable via the TF output metric of interest *Forecasted Number of Institutions Connected by the NREN (Y_6)*, revealed that the Cross-validated Communality (H^2) tested positively, while the Cross-validated Redundancy (F^2) tested negatively. Hence, the example NREN model instantiation's measurement indicators are well-suited to forecasting the future NREN reach, but the defined structural relationships are not well-suited for this.

C. Evaluation of the Research Propositions using the SEM Regression Results

Using the calculated path coefficients in Table 3 and the path coefficient significance test results in Table 5, the research propositions defined for the example NREN model instantiation in Section III.B were evaluated as follows:

- **Research Proposition H1:** While the path coefficient of γ_1 = 0.1659 supports the direction of the proposed relationship between *Government Influence of the NREN (ξ_1)* and *NREN Infrastructure Orientation (η_1)*, the path coefficient was judged to not be significant at the maximum allowed significance level of α = 0.10. Hence, this hypothesized relationship was rejected.
- **Research Proposition H2:** This hypothesized relationship between *NREN Infrastructure Orientation (η_1)* and *NREN Advanced Services Capability (η_2)* was not rejected, since the path coefficient β_1 = 0.5097 supported the direction of the proposed relationship and the path coefficient was judged to be significant at the maximum allowed significance level of α = 0.10.
- **Research Proposition H3:** Since the path coefficient γ_2 = -0.2073 did not support the direction of the hypothisized relationship between *Government Influence of the NREN (ξ_1)* and *NREN Advanced Services Capability (η_2)*, this research proposition was rejected.
- **Research Proposition H4:** The postulated relationship between *NREN Advanced Services Capability (η_2)* and *Current NREN Reach (η_3)* was rejected since the path coefficient β_2 = 0.0704 was judged to not be significant at the maximum allowed significance level of α = 0.10.
- **Research Proposition H5:** This research proposition was rejected, since the path coefficient β_3 = -0.0095 did not support the direction of the hypothesized relationship between *NREN Advanced Services Capability (η_2)* and *Forecasted NREN Reach (η_4)*.
- **Research Proposition H6:** Since the path coefficient γ_3 = 0.1586 was judged to be significant at the maximum allowed significance level of α = 0.10 and supported the direction of the proposed relationship between *Scope of the NREN Mandate (ξ_2)* and *Current NREN Reach (η_3)*, this research proposition was not rejected.
- **Research Proposition H7:** This research proposition, which hypothesized a relationship between *Scope of the NREN Mandate (ξ_2)* and *Forecasted NREN Reach (η_4)*, was rejected on the basis of the path coefficient γ_4 = -0.0229 not supporting the direction of this relationship.
- **Research Proposition H8:** Since the path coefficient β_4 = 0.9964 was deemed to be significant maximum allowed significance level of α = 0.10 and also supported the direction of the hypothesized relationship between *Current NREN Reach (η_3)* and *Forecasted NREN Reach (η_4)*, this research proposition was not rejected.

V. CONCLUSIONS

This paper presented an example model instantiation of Staphorst, Pretorius and Pretorius' framework for SEM based DF for TF [14][15] for the NREN technology domain. The example NREN model instantiation was created using deductive reasoning from knowledge gained through action research in SANReN, as well as data captured by TERENA in its yearly NREN compendium series [2][1]. Data from this compendium was also used to perform a PLS regression analysis to not only determine the path coefficients and indicator loadings in the example NREN model instantiation, but also its reliability and validity. Using the path coefficient results a number of research propositions related to hypothesized relationships in the example NREN model instantiation were tested.

From the PLS regression results obtained for the example NREN model instantiation it can be concluded that all of the technology related indicators, with the exception of *Access to IPv6 Addresses (Y_4)*, were able to adequately measure their respective technology related model constructs. Similarly, all context related indicators adequately measured their associated context related constructs. Furthermore, results showed that several postulated relationships between technology related constructs (such as the relationship between *NREN Infrastructure Orientation (η_1)* and *NREN Advanced Services Capability (η_2)*), as well as relationships between technology and context related constructs (such as the relationship between *Scope of the NREN Mandate (ξ_2)* and *Current NREN Reach (η_3)*), were supported by the secondary data published in the TERENA NREN compendium series [2][1]. There were, however, several postulated relationships (such as the relationship between *Scope of the NREN Mandate (ξ_2)* and *Forecasted NREN Reach (η_4)*) that were rejected by this data. Also, during the reliability and validity evaluation of the example NREN model instantiation it was discovered that, while the measurement portion of the model was capable of contributing adequately to the forecasting of the future NREN reach, the same was not true for the structural portion of model.

In order to address these weaknesses of this paper's example NREN model instantiation, which was derived through deductive reasoning, future research will entail a qualitative study [20] that will attempt to identify improved endogenous and exogenous model constructs, technology indicators and interactions between the various indicators and constructs. The unit of analysis [20] for this qualitative study will be an NREN, while the population will be all NRENs in existence worldwide at the time of the study. Data collection will be accomplished through online surveys with open-ended questions as data collection instrument [2]. Respondents will be selected from the global community of NREN specialists through a snowball sampling approach [20]. Sufficiency of the sample size will be determined through the principle of data saturation [13]. Analysis of the collected qualitative data will firstly entail narrative inquiry by means of a process of theme extraction [13]. Thereafter frequency analysis will be performed on the extracted themes in order to produce a final set of importance ranked indicators, constructs and interconnections from which the NREN model instantiation will be constructed [13]. Testing the reliability and validity of the collected qualitative data will be accomplished by means of theory triangulation [13], as well as data triangulation [13] using as baseline published technology indicators from secondary data sources, such as TERENA's NREN compendium series. A quantitative study will then be performed to determine, through PLS regression, the indicator loadings and path coefficients of the NREN model constructed during the qualitative study. As with the qualitative study the population will be all NRENs in existence at that point in time, with the unit of analysis being a single NREN [20]. Quantitative online surveys, constructed using close-ended questions with Likert scaling, will be used as data collection instrument [20]. Senior managers at all of the NRENs in the population will be selected through a process of convenience sampling [20] as respondents for these surveys.

REFERENCES

[1] "2011 TERENA compendium of national research and education networks in Europe," Retrieved 30 December 2013 World Wide Web, http://www.terena.org/activities/compendium/2011/pdf/TER-C11-complete-web.pdf, 2011.

[2] "2012 TERENA compendium of national research and education networks in Europe," Retrieved 12 January 2013 World Wide Web, http://www.terena.org/activities/compendium/2012/pdf/TER-C12-final-web.pdf, 2012.

[3] "The South African national research network (SANReN)," Retrieved 27 December 2013 World Wide Web, http://www.sanren.ac.za/, 2013.

[4] Chang, H., J. Gausemeier, S. Ihmels and C. Wenzelmann; "Innovative technology management system with bibliometrics in the context of technology intelligence," *Trends in Intelligent Systems and Computer Engineering: Lecture Notes in Electrical Engineering*, vol. 6, pp. 349-361, 2008.

[5] Grupp, H.; *Foundations of the Economics of Innovation: Theory, Measurement and Practice.* Cheltenham: Edward Elgar, 1998.

[6] Gustavsen, B., "Action research, practical challenges and the formation of theory". *Action Research*, vol. 6, no. 4, pp. 421-437, 2008.

[7] Lichtenthaler, E.; "Technological change and the technology intelligence process: a case study," *Journal of Engineering and Technology Management*, vol. 21, pp. 331-348, 2004.

[8] Nyberg, A. and S. Plamgren; "Using indicators for technology monitoring: steps toward a proposed framework," *Chalmers University of Technology, Master of Science Thesis in the Master Degree Programme: Business Design*, 2011.

[9] Porter, A.L.; "QTIP: quick technology intelligence processes," *Technological Forecasting and Social Change*, vol. 72, pp. 1070–1081, 2005.

[10] Porter, A.L.; "Tech mining to drive open innovation," in *Proceedings, First International Conference on Technology Innovation, Risk Management and Supply Chain Management*, pp. 1-13, 2007.

[11] Ringle, C.M., S. Wende and S. Will; "SmartPLS 2.0 (M3) beta," Retrieved 24 January 2014 World Wide Web, http://www.smartpls.de, 2005.

[12] Sohn, S.Y. and T.H. Moon; "Structural equation model for predicting technology commercialization success index (TCSI)," *Technological Forecasting and Social Change*, vol. 70, no. 9, pp. 885-899, 2003.

[13] Staphorst, L.; "The impact of intellectual property rights from publicly financed research and development on governance mode decisions for research alliances," *Gordon Institute of Business Science, Master of Business Administration Thesis*, 2010.

[14] Staphorst, L., L. Pretorius and M.W. Pretorius; "Structural equation modelling based data fusion for technology forecasting: a generic framework," in *Proceedings of PICMET2013*, July 2013.

[15] Staphorst, L., L. Pretorius and M.W. Pretorius; "Structural equation modelling based data fusion for technology forecasting: framework development and example in the national research and education network technology domain," *Technological Forecasting and Social Change: PICMET2013 Special Edition*, in press.

[16] Steinberg, A.N.; "Context-sensitive data fusion using structural equation modeling," in *Proceedings, IEEE 12th International Conference on Information Fusion*, July 2009.

[17] Steinberg, A.N. and G. Rogova; "Situation and context in data fusion and natural language understanding," in *Proceedings, IEEE 11th International Conference on Information*, 2008.

[18] Wald, L.; "Some terms of reference in data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, May 1999.

[19] Vinzi, V.E., W.W. Chin, J. Henseler and H. Wang, H.; *Handbook of Partial Least Squares: Concepts, Methods and Applications*. Berlin, Heidelberg: Springer, 2010.

[20] Zikmund, W.; *Business Research Methods, 7th Edition.* Ohio: South-Western Cengage Learning, 2003.