Combination of probabilistic and deterministic models in degradation prognostics with limited data

Petteri Ojala
Materials Science, Faculty of Engineering and Natural Sciences, Tampere University, Finland, E-mail: petteri.ojala@tuni.fi

Jari Rämö
Materials Science, Faculty of Engineering and Natural Sciences, Tampere University, Finland, E-mail: jari.ramo@tuni.fi

Minna Niittymäki
Electrical Engineering, Faculty of Information Technology and Communication Sciences, Tampere University, Finland, E-mail: minna.niittymaki@tuni.fi

Juha Miettinen
Materials Science, Faculty of Engineering and Natural Sciences, Tampere University, Finland, E-mail: juha.miettinen@tuni.fi

Sufficient and high quality data is a requirement for accurate modeling. Modern data acquisition practices and technologies provide good tools to meet these demands. Regardless of carefully implemented experiments, the collected data might not be detailed enough. Sometimes acquired data reveals phenomena that should be studied with a different scale or resolution, but used set-up cannot provide required data. In an ideal case, test set-up is updated to provide required capabilities and new experiments are performed. Often limited resources prevent revision of equipment and new tests, thus other solution has to be found. Lacking data can be supplemented with other existing information, knowledge, and new specific data acquisition, if additional research is possible. This paper concentrates on this problematic field.

This study focuses on modeling of CAN-bus connector degradation based on limited data and improving usefulness of the existing data with supplementary information. The supplementary data was gathered with specific measurements and varying of the initial test procedure. In interpretation of the data, the knowledge of the design of CAN-bus connector and physics behind measured quantity were used. Supplementary information was used to rule out otherwise plausible model options. Acquired information is collected into the model that comprises deterministic parts and statistical components. The model can be applied in prognostics of CAN-bus, and used methods give tools to work with processes that are partially unknown.

Keywords: Degradation, model, deterministic, probabilistic, prognostics.

1. Introduction

In optimal situation, fundamental properties of a device under test (DUT) are well-known. In addition, methods and a goal are well-defined. Then specification of experiments, equipment etc. can be performed accordingly. This might be possible in development of new versions of existing products when there is lots of *a priori* knowledge of similar objects in version history.

In basic research or the development of new methods, there is not such a knowledge available. This leads easily to a trial-and-error process where methodology, equipment and goal are updated according to new observation after each iteration round. The iterative method demands lots of time and other resources, and eventually iterative development is stopped when sufficient result is achieved or allocated resources are spent.

However, it is possible to enhance usability of already acquired information with supplementary information. Additional measurements and other available information can be used in the interpretation of already existing information without the extensive use of resources.

2. Theory

Modeling and simulation are basic tools in many fields of science. Especially the branches of science whose data is in numeric form by default or data can be presented in numeric form through operationalization has found modelling and simulation very efficient tools. In operationalization (Weaver 2015) indicating factors of studied phenomena are described by measurable variables. For example, survey data that has qualitative structure is transformed into quantitative that makes use of statistical methods possible.

Very large amount of data can be compressed with models and described with a few statistical parameters. Above all, models makes prediction of the future development possible, but also backward extrapolation (Gilad 2017) to explain the history of events prior to observed moment.

Each scientific discipline has developed models to fit their needs. There are many ways to cluster different model types, but in general there is two main classes of models: probabilistic (stochastic) and deterministic models (Ljung 1994). A typical example of a stochastic model is Brownian motion (Ross 1996) where statistical properties of process are known, but the exact development of a process in a next time step is random. Deterministic models are based on mathematical functions and are predictable from step to step.

Stochastic models are typically applied to complex systems where is many interacting components and factors that are random by nature. For example, economics use stochastic models (Shreve 2004) to simulate processes that include many affecting factors such as laws of demand and supply, speculation of political systems, international tension, psychology of human behavior etc. that are

Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference
Edited by Piero Baraldi, Francesco Di Maio and Enrico Zio

ISBN: 978-981-14-8593-0; doi:10.3850/978-981-14-8593-0_3912-cd

impossible to predict with deterministic models. In engineering, stochastic models are used in simulation of complex, interacting systems with inherent randomness (Borrie 1992). Deterministic models are suitable for well-known, relatively simple systems such as electronic circuits and ballistic trajectories that are based on elementary physics.

Prerequisite for successful data based modelling and simulation is data that describes studied subject with relevant accuracy. Sampling frequency of data acquisition must be high enough to detect rapid phenomena, dynamic range and sensitivity of a gauge must be able to observe small changes in the measured quantity within the dynamic range of the data acquisition equipment.

However, even the best-planned test procedure can fall short of these demands, if conditions change unexpectedly or new phenomenon is observed outside the specifications of used test set-up. In that case, other methods can be used to fill in the cap in data. Well-known solutions to this problem are statistical methods such as Bayesian approach or methods concerning censored or truncated data (Meeker 1998).

As an addition to strictly mathematical methods, all information already available and easily achievable can be applied. Some specific measurement data or other information of phenomenon can be used rule out otherwise plausible options and extract information out of existing but sparse data.

3. Initial goal and available data

3.1 Initial experiments and test set-up

Initial goal was to study the waterproofness of CAN-bus connector component under continuous salt fog exposure following standard SFS-EN ISO 9227. In CAN-bus connector component is five pin-socket pairs that transfer signal from cable to cable. Four wires are for signal and one for grounding. Resistance between pin-socket pairs of different signal wires was selected as an observed quantity, and as an indicator of degradation. In total, there is ten observed resistance signal channels in each connector component sample.

When space between pins is dry, resistance is very high and out of the dynamic range of normal multimeters. After salt water starts to seep through into connector structure, resistance decrease into dynamic range of the used multimeter. This threshold was used as an indicator of waterproofness failure, yielding also failure time.

The failure times vary from sample to sample, and it was not practical to remove failed samples, so test run was continued until all samples had failed. This yielded more detailed data of the development of the degradation process as a byproduct. Also, test run had to be paused from time to time that yielded useful information about possible degradation process (Chapter 4.2). Studied connector component with relevant details is presented in Figure 8 (Chapter 4.3).

3.2 Censored data and dead time

Beginning part of the degradation data is above the upper limit of dynamic range of the multimeter. This kind of data is called top-coded (OECD 2007) or right censored (Nelson 2004) in statistics, and it is only known that values are higher that threshold value i.e. the upper limit (*UL*) in dynamic range of the multimeter. Resistance values above *UL* are set to constant value (150 MΩ). This condition is similar to the concept of dead time (Gilad 2017) in data acquisition. Graph of typical development of resistance between pins is presented in figure 1. Directly useful data is between time markers *Tt* and T_{end}, that mark dropping of resistance values to the dynamic range of the multimeter and ending of the data acquisition, respectively.

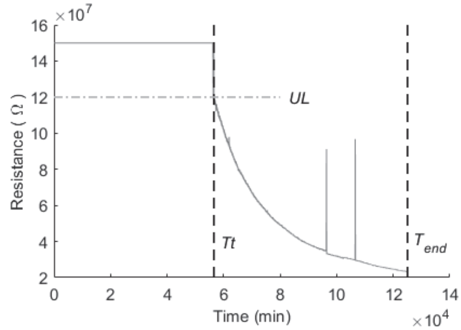

Fig. 1. Censored resistance data with upper limit of dynamic range *UL* and limits of available data *Tt* and T_{end}.

3.3 Transition time

A useful concept in analysis is transition time *Tt* (Ojala 2017). This is time when resistance values has dropped constantly below *UL*. It has important role in the modeling of degradation process. The shape and endpoint of the resistance graph depend clearly on *Tt*, thus it was selected as a parameter of the degradation model (Chapter 5.4). *Tt* is similar to the concept of first hitting time (Redner 2001) that is used in stochastic processes. Resistance graphs of signal channels with different *Tt* values are presented in Figure 2. Endpoint resistance values as a function of time are presented in Figure 3.

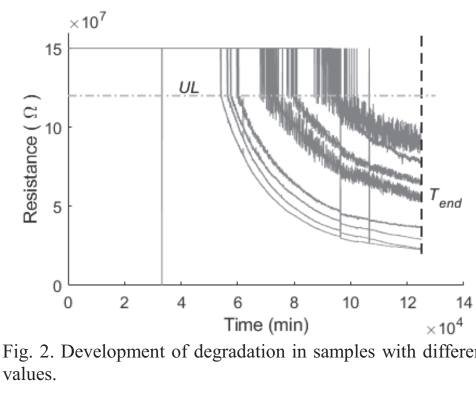

Fig. 2. Development of degradation in samples with different *Tt* values.

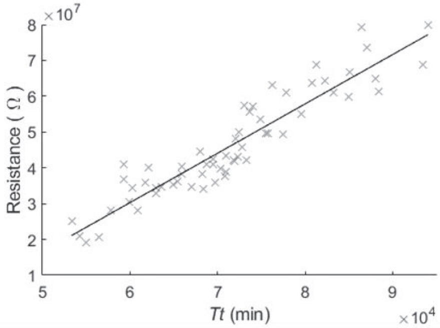

Fig. 3. Dependence of endpoint resistance value as a function of *Tt* time.

3.4 Models

According to the nature and physics of observed quantity, it is well-founded assumption that mechanism of degradation is the diffusion of humidity. Diffusion is modelled according to Fick's laws (Smith 1993). Solution to basic diffusion equations are often exponential functions:

$$F_E(t) = Ae^{-Bt} + E \tag{1}$$

where t is time, A, B and E are constants depending on a process.

Other model option is also based on diffusion, but it uses more information about structure of the DUT. In figure 4 is shown schematic drawing of connector end and diffusing salt water.

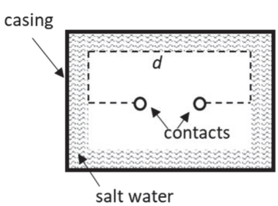

Fig. 4. Schematic drawing of connector face and conductive path between two contact pins/sockets along diffusion edge.

Between the connector parts is a narrow space where salt water diffuses. Salt water forms a connective path along diffusion edge, and total resistance along path *d* between the contact points of two signal wires decreases to the dynamic range of the multimeter. As a diffusion edge proceeds, resistance value decreases as a function of time. Solution to differential equation describing this process follow power functions:

$$F_P(t) = At^B + E \tag{2}$$

where t is time, A, B and E are constants depending on a process.

4. Supplementary information

Information that were used to help modeling can be divided into two categories: measurements and other information.

4.1 *Extra measurements*

Data described in Chapter 3.2 is censored from both left and right. The beginning and the end values of resistance are unknown. These data points are crucial in model selection. There is many plausible options that fit well to existing data, but with two extreme values, it is easier to assess different options and rule out unsuitable ones.

4.1.1 *Intact condition*

Resistances of the intact samples were measured with a test setup, which is typically utilized in the resistivity measurements of insulation materials. The measurements were made in accordance with volume resistivity standard of insulation materials IEC 60093 when it was applicable. The resistance measurements were made using high sensitivity Keithley 6517B electrometer. The measuring voltage was 50 V. The test voltage was maintained for 60 s in order to reach the steady state value of current. The resistance was defined from the average current value in the end of the measurement period.

Although accuracy of extremely high resistance values is not very good, it can be used in the assessment of goodness of models. Histogram of resistance values and fitted lognormal distribution are shown in Figure 5.

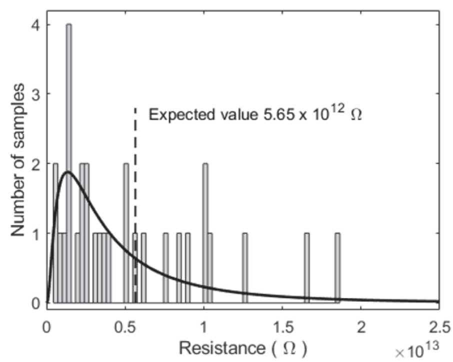

Fig. 5. Histogram of intact samples' resistance data and fitted lognormal distribution with expected value.

4.1.2 *Fully saturated condition*

The endpoint of the degradation process was approximated by dipping open connector to salt water and closed before measurements to simulate fully saturated condition where salt water has filled space between connector plug and socket parts. Histogram of resistance values and fitted lognormal distribution are shown in Figure 6.

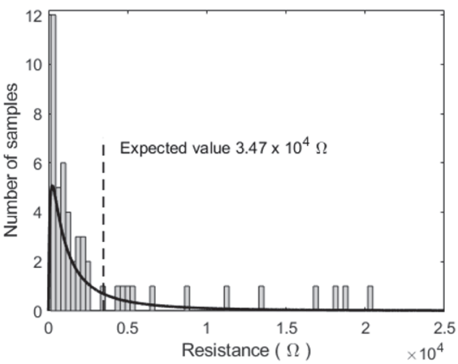

Fig. 6. Histogram of saturated samples' resistance data and fitted lognormal distribution with expected value.

4.2 Variation in test procedure

The initial procedure did not include any pauses in testing. The goal was to run salt fog test until every monitored channel had reached the failure criterion. During the salt fog test, there was a need to pause the test from time to time for other tests using the same equipment simultaneously. These pauses caused rapid changes to the resistance readings.

This phenomenon was studied with the systematic procedure including drying phases amidst continuous salt fog exposure (Ojala 2019). Both beginning of drying and continuing of humidity exposure after drying had rapid effect on the resistance compared with the development of the long term degradation process. Typical development of the resistance is presented in Figure 7. Beginning and end of the pause are marked with dash lines and *UL* with dash-dot line.

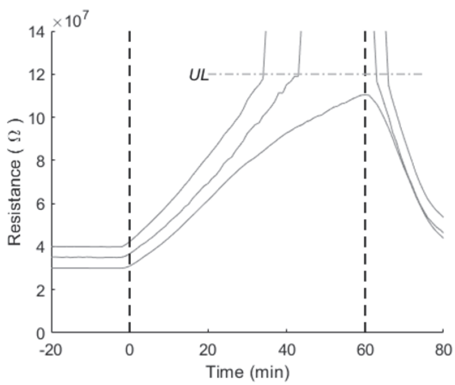

Fig. 7. Three example graphs of resistance responses to salt fog exposure pause.

4.3 *Design of DUT*

Design of the studied CAN-bus connector component is a typical extension cord design. Each signal wire is connected to a pin on one side and to a socket on other part. When parts are pressed together pin and socket form a galvanic contact enabling signal transfer. Parts are overlapping with silicon flanges for waterproofness between casing parts. DUT is presented in Figure 8.

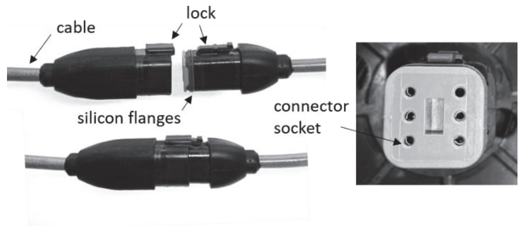

Fig. 8. Open (left top) and closed (left bottom) CAN-bus connector and socket side of connector pair (right).

5. Results and discussion

The development of the model follows deductive process that rules out some options and forces plausible within boundaries. Components of the deduction with clustering to four topic areas are presented in Figure 9.

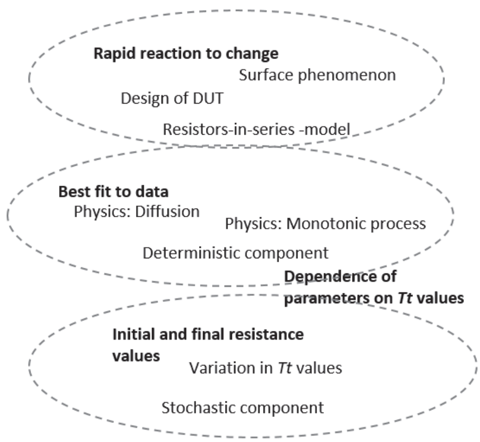

Fig. 9. Components of deduction process and four topic areas (in bold).

5.1 Rapid reaction to change

Variation in the test procedure narrowed selection of possible models. Rapid changes in the resistance suggests that changes in conductivity must happen on a surface of the DUT. An equivalent electric circuit that fits this information and good fit of exponential function with two exponential components can be formed with three resistors. An equivalent circuit is presented in Figure 10.

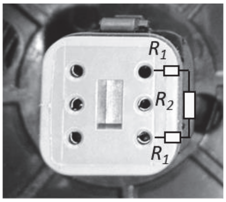

Fig. 10. Circuit model of observed pin-to-pin resistance.

Resistance measurement from pin-to-pin with normal multimeters or data acquisition equipment requires a conductive path between pins. The path is formed of three pieces: pin-to-edge, along-edge, edge-to-pin. Resistances between pin and edge and along edge are modeled with resistors R_1 and R_2, respectively.

Diffusion of salt water follows exponential function and resistance drops accordingly, so the resistance of both resistors follows exponential function as a function of time, and total resistance of resistors in series follows Eq. (3) (Chapter 5.2).

Formation of the closed circuit requires diffusion to penetrate waterproof flanges, reach two contacts and wet the edge and flanges. This process takes time that is indicated by *Tt*. However, the conductive path along edge is easily broken when salt fog exposure is paused and sample dried, and mutually quickly restored when exposure is continued as can be seen in Figure 7.

5.2 Best fit to data

In selection of a basic model for the degradation both exponential and power function yielded good fit to available data. Fit for each signal channel was made to the uncensored part of the data between signal channel specific time *Tt* and the end of data acquisition t_{end}.

Fit to function Eq. (1) that has one exponential component was had flaw even statistical fitting parameters were good. At the end of data there was small, but notable difference between the fit and the data. This deficiency was corrected by adding a second exponential term yielding three-component model:

$$F_{E2}(t) = Ae^{-Bt} + Ce^{-Dt} + E \qquad (3)$$

where *t* is time, *A*, *B*, *C* and *D* are parameters of the process, and *E* is constant describing endpoint of diffusion process. Fit to this model was very good in whole area of observation. Need for two exponent components points towards two simultaneously progressing degradation processes that affect to total resistance.

Rapid reaction to humidity condition change also supports exponential function model over power function model by explaining observed reactions so that two-component exponent function is more suitable option for the model.

5.3 Initial and final resistance values

The initial resistance values of intact sample components (Chapter 4.1.1) were so high, expected value 5.65 GΩ, that neither exponential nor power functions fits did not reach plausible level in backward extrapolation, thus degradation process cannot be continuous. Plausible mechanism could be diffusion over structural threshold resulting in drastic drop of resistance. After this threshold, the process develops according to the data.

Fully saturated condition (Chapter 4.1.2) gives the endpoint of degradation process i.e. value of constant parameter *E* in Eq. (3). Expected value of *E* is 34.7 kΩ that is negligible in comparison with the magnitude of the acquired data, and it does not affect fitting results or the simulation results.

5.4 Dependence of parameters on *Tt* values

Transition time *Tt* has major impact on the development of the process and the curve shape (Chapter 3.3). Sooner the *UL* level is reached, steeper is the degradation process. When data is fitted to exponential model Eq. (3), observation is consistent with the parameters of fitted curves.

Parameters *A* and *C* are dependent by the nature of data and the model. At time *Tt* value of *R* is fixed to *UL,* and when *A* is known $C = (UL - A)$. Parameters *A* and *D* have linear dependence on *Tt*, but parameter *B* has no correlation with *Tt*. However, *B* has correlation with *D*, so *B* can be expressed as a function of *D*. Linear functions that describe these dependencies are of form:

$$L(x) = p_1 x + p_2 \qquad (4)$$

where *L* is degradation model parameter *A*, *D* or *B* and parameters p_1 and p_2 are specific to that degradation model parameter, and *x* is *Tt* for *A* and *D* and *D*(*Tt*) for *B*. Even parameters *B* and *D* are interdependent, both are kept in notation for clarity reasons.

Values of parameters p_1 and p_2 with 95 % confidence limits in brackets for each degradation model parameter *A, D,* and *B* are presented in Table 1.

Table 1. Parameter values of linear models of degradation model parameters.

	p_1	p_2
A	-1110 (-1480 ; -748)	1.40×10^8 $(1.14; 1.66) \times 10^8$
D	1.40×10^{-10} $(0.40 ; 2.40) \times 10^{-10}$	$-1.41 \cdot 10^{-5}$ $(-2.12 ; -0.70) \times 10^{-5}$
B	2.16 (1.10 ; 3.22)	$-4.90 \cdot 10^{-5}$ $(-5.48; -4.32) \times 10^{-5}$

Development of degradation parameters *A* and *D* as a function of *Tt* and development of parameter *B*(*D*) as a function of *D*(*Tt*) are presented in Figures 11, 12 and 13, respectively.

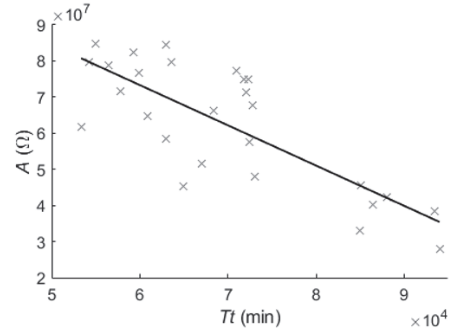

Fig. 11. Parameter A as a function of Tt, and regression line.

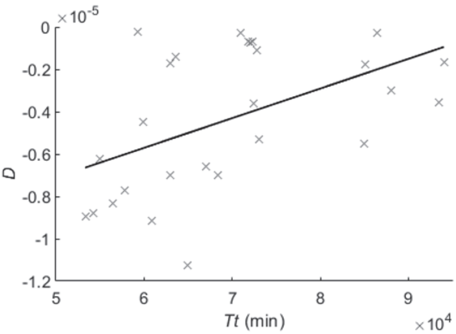

Fig. 12. Parameter D as a function of Tt, and regression line.

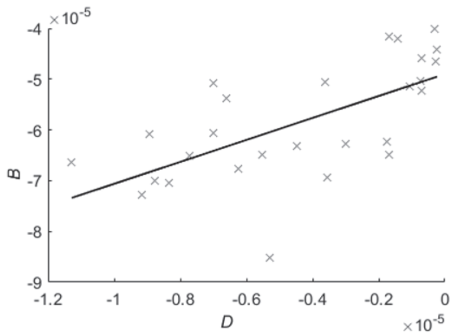

Fig. 13. Scatter plot of parameters D and B, and regression line.

Degradation model as a function of t and Tt, is formed by inserting linear models Eq. (4) to basic model Eq. (3), and replacing t with τ that is time starting from Tt yielding:

$$R(\tau, Tt) = A(Tt)e^{-B(D(Tt))\tau} + (UL - A(Tt))e^{-D(Tt)\tau} + E \quad (5)$$

According to the properties of the data, all resistance values prior to the Tt censored are censored and set to fixed value 150 MΩ.

Development of monitored resistance values follow deterministic model in general, but as can be seen in Figures 11, 12, and 13 there is variation that is not covered by the linear model. To take into account this feature, the model is supplemented with probabilistic factors. Deterministic parameters A and D are made random variables with normal distributed values: $A \sim N(A(Tt), \sigma_A)$ and $D \sim N(D(Tt), \sigma_D)$. Same approach is used to transition time: $Tt \sim N(\mu_{Tt}, \sigma_{Tt})$. Parameters of the distribution are computed from available data.

Due to the nature of diffusion and used principle of modeling that assumes monotonic, decreasing behavior of resistance values of degradation model parameters are limited. If the value of D or B exceeds value zero and becomes positive, it is set to fixed value zero.

5.5 Simulation of degradation

Simulations uses Tt distribution as an input. Each realization is based on a Tt value generated from distribution. With fixed Tt parameters A, B, and D are generated and inserted into the deterministic model presented in Eq. (3), yielding simulated development of the resistance after time Tt.

Two sets of simulations were generated. One using Tt distribution $Tt \sim N(\mu_{Tt}, \sigma_{Tt})$ and other with $Tt_{ref} \sim N(\mu_{Ttref}, \sigma_{Ttref})$ where parameters are calculated from the reference data. The reference data is similar to the data used in the modeling but acquired in separate test runs. In total 3000 realizations were generated with each distribution. In Figure 14 is presented 45 randomly selected realizations generated from $N(\mu_{Tt}, \sigma_{Tt})$ distribution.

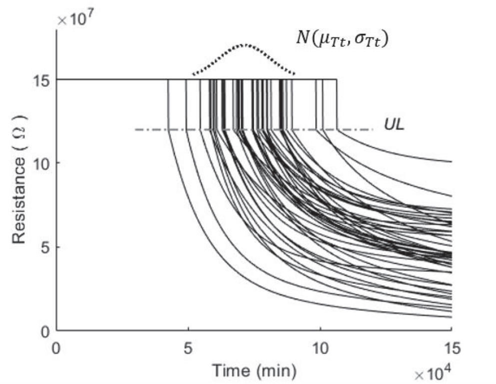

Fig. 14. Degradation simulation with Tt distribution data used in modelling. Only 45 out of 3000 realizations in presented.

Verification of the developed model Eq. (5) is done by comparing endpoint distributions of the data and the simulation. In the validation, the endpoint distribution of the reference data and simulations using Tt distribution of the reference data were applied. Endpoint distributions of data (*Data*, black dash-dot line), simulations (*Sim.* , black solid line), reference simulation (*Ref sim.* , gray dash-dot line), and reference data (*Ref. data*, gray solid line) are presented in Figure 15, and corresponding parameters in Table 2.

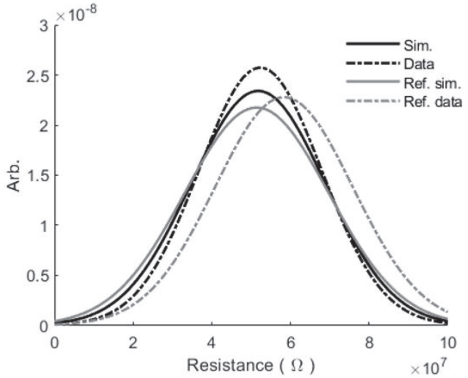

Fig. 15. Circuit model of observed pin-to-pin resistance.

Table 2. Parameters of endpoint distributions.

Resistance (10^7 Ω)	μ	σ
Data, model	5.83	1.66
Simulation	5.86	1.79
Data, reference	5.84	1.75
Simulation reference	5.12	1.83

Assessment of the validation and the verification quality is based on the parameter values and an overlapping percentage of distributions (common area of two probability distributions functions). Parameters of the endpoint distributions of data and simulation are very close to each other. Overlapping percentage is 95.9 % that verifies modelling.

Validation results are not so good. Variation of the reference simulation is 4.57 % larger that variation of reference data. Mean value of the reference simulation is 12.3 % larger that mean value of the reference data. Overlapping percentage is 84.3 %. Verification result is not optimal, but considering deficits in the available data, it is sufficient.

6. Conclusions

Presented case study showed how usefulness of the limited degradation data of CAN-bus connector was enhanced with supplementary information. Specific measurements, basic information of physics behind observed phenomenon, and deduction process gave credibility to the developed model. Acquisition of the supplementary information can be very useful option compared to new test run and equipment updates, if resources are sparse.

Extra measurements and supplementary information was found valuable in the interpretation of the data that was acquired with the equipment planned for other task. With supplementary information, it was possible to develop the hybrid model combining probabilistic and deterministic components that can be used in the prognostics of CAN-bus connector. Also, analysis revealed the significance of transition time *Tt* in description of the process. It determinates development of the degradation process, and as itself, it could be used as indicator of failure time or failure criterion.

Collecting of information related to the original data and conditions from many different aspects that are presented in Figure 9 revealed useful details. Visual inspection of the data graphs can give insight of data analysis methods that could be applied. *A priori* information is always available at some level. In this case, assumption of diffusion process and simple resistor circuit model were a basis of the deduction process.

Knowledge of the initial and the final resistance values suggested that the deterministic model is not sufficient, but it should be complemented with other factor. Since the available data did not reveal details of the development of the resistance between the starting of the test and the transition time *Tt*, censored part of development was replaced with the probabilistic model of *Tt* that summarizes censored part of the process.

Especially variation in the test procedure yielded important information. In this case study, rapid response to changing conditions helped in deduction of mechanism and model selection. In general, gradients in environmental conditions cause significant stress to structures and components. For instance freezing and melting of water causes erosion and deterioration in building structures. This effect is well known and temperature cycling is commonly used method in testing. However, test set-ups and equipment including multiple stress factors and cycling of each parameter becomes easily complicated and very expensive. Therefore, use of limited supplemental information can be a viable option if resources do not allow detailed study of every factor.

Presented results give starting point for the future research. Studied modeling concerns only the development of resistance between the signal pins that is interpreted as an indicator of degradation. Even in fully saturated condition, resistance is still in range of kilo-ohms that is high enough for undisturbed signal transfer in the CAN-bus. But, water inside a connector very likely corrodes pins and wires, and contacts between them. Durability of corroded structures reduce and make them more susceptible to a mechanical damage due vibration that is a common stress factor in components used in vehicles and work machinery. Effect of material degradation would be important addition to the modeling of degradation process.

Acknowledgement

The authors gratefully acknowledge Dunderberg foundation (*K.F ja Maria Dunderbergin testamenttisäätiö* in Finnish), Business Finland and the companies Valtra Oy, Sandvik Mining and Construction Oy, TVO Oyj and Fortum Power and Heat Oy for funding the research.

References

Borrie, J. (1992). *Stochastic Systems for Engineers. Modelling, Estimation & Control*. Prentice Hall International Ltd. UK.

Gilad, E., Dubi, B., Geslot, B., and Kolin, A. (2017). Dead time corrections using the backward extrapolation method. *Nuclear Inst. and Methods in Physics Research, A*, 854, 53-60.

Ljung, L., and Glad, T. (1994). *Modeling of dynamic systems*. Prentice Hall Inc. New Jersey, USA.

Meeker, W.Q. & Escobar L.A. (1998). *Statistical Methods for Reliability Data*. New York: John Wiley & Sons, Inc.

Nelson, W. (1990). *Accelerated Testing, Statistical Models, Test Plans, and Data Analyses*. New York: John Wiley & Sons, Inc. USA.

OECD, *Glossary of statistical terms* (2017). Web page: https://stats.oecd .org /glossary/detail.asp?ID=7011 (accessed 14.1.2020).

Ojala, P., Hietala, J-P., Miettinen, J., Julkunen, P., Nieminen, I., (2017), Modelling of seep through of humidity to electric connector with stochastic processes, In Cepin, M., Bris, R. (Eds.) *Safety & reliability: theory and applications*, Proceedings of the 27th European Safety and Reliability Conference, ESREL 2017, Portorož, Slovenia. CRC Press.

Ojala, P., Rämö, J., Nieminen, I., Miettinen, J., (2019), Modelling of seep through of humidity to electric connector with stochastic processes, In Beer, M., Zio, E. (Eds.) *Safety & reliability: theory and applications*, Proceedings of the 29th European Safety and Reliability Conference, ESREL 2019, Hannover, Germany. Research Publishing, Singapore.

Radner, S. (2001). *A Guide to First-Passage Processes*. Cambridge University Press, UK.

Ross, S. (1996). *Stochastic Processes 2nd ed*. New York: John Wiley & Sons, Inc.

Shreve, S. (2004). *Stochastic Calculus for Finance II.* Springer-Verlag New York, USA.

Smith, W.F. (1993). *Foundations of materials science and engineering 3rd ed.* McGraw-Hill, Inc. USA.

Weaver, S., (2015). Operationalization, In Wolfgang Donsbach (Ed.) *The International Encyclopedia of Communication 1st ed*. John Wiley & Sons, Ltd. Web page: https://onlinelibrary .wiley.com/doi/pdf/10.1002/9781405186407.wbieco01 2.pub3 (accessed 18.1.2020).