HydroCen brief Nr. 14

Forprosjekt - HighFlex

FORFATTER	Espen Eberg
PROSJEKT I HYDROCEN	WP 2, 5.2.10
DATO	2023-05-09
SAMARBEIDSPARTNERE	SINTEF Energi, NTNU

Formål

Forprosjektet HighFlex skal konkretisere prosjektinnhold i form av hypoteserer, FoU-utfordringer, prosjektstruktur og prosjektdeltakere i prosjektforslaget HighFlex inn mot søknad om KSP-prosjekt i februar 2022. HighFlex bygger videre på funn og forslag til videre arbeid i KPN-prosjektet HydroStator som ble avsluttet i 2021. Som en del av arbeidet skal også måleresultater fra HydroStator og HydroCen publiseres og presenteres på konferansen NORD-IS'22. Dette vil synliggjøre relevant forskning og styrke en fremtidig KSP-søknad.

Vedlegg

1. Artikkel, *Multi-stress Cyclic Testing of Roebel Stator Bars for Hydropower*
2. Artikkel, *Use of Data-Driven Approaches for Defect Classification in Stator Winding Insulation*
3. Prosjektforslag, *KSP HighFlex*

Bakgrunn

Intermitterende drift av vannkraftanlegg gir nye/økte påkjenninger, og det er behov for mer kunnskap om hvordan eksisterende og nye vannkraftanlegg tåler slike påkjenninger. I KPN-prosjektet HydroStator undersøkte vi tilstanden til høyspenningsisolasjonen i en generator som hadde vært i drift i ca. 50 år. Det ble funnet at tilstanden til det aktuelle isolasjonssystemet var svært god. Videre studier med akselerert multistress aldring, hvor store lastvariasjoner med påtrykt samtidig som generatorspolene var spenningssatt, viste betydelig økt PD-aktivitet når temperaturen varierte. Dette ble tilskrevet termisk ekspansjon og sammentrekning av isolasjonsystemet. Ved langvarig intermitterende drift vil dette kunne påvirke levetid og pålitelighet negativt.

Basert på dette har SINTEF og NTNU jobbet med et prosjektforslag mot vannkraftbransjen for å videreføre studiene. Følgende aktiviteter har vært utført i forprosjekt HighFlex

1. Utarbeiding av prosjektforslag og møter med mulige partnere
2. Publisering av arbeid på lastvariasjon
3. Publisering av arbeid på datadrevne tilstandskontrollmetoder (i samarbeid med HydroCen-prosjektet TwinLab)

Resultater

Prosjektforslag HighFlex

Det ble utformet et prosjektforslag hvor formålet er å utvikle kunnskap og metoder for å forstå og predikere endringer i helsetilstanden til generatorer under intermitterende/ dynamisk drift. Arbeidet vil bestå av akselerert lastsykling av relevante komponenter i generatorriggen som bygges som en del av HydroCen Labs. Tilstandsprediksjon vil baseres på fysikalske, hybride og datadrevne metoder. For sistnevnte vil transparente og forklarbare metoder (*explainable AI*) bli benyttet. Videre er det lagt opp til metodeutvikling innenfor trefase PD-måling og dielektriske målinger. Om potensielle prosjektdeltakere har andre metoder de mener har potensiale til å utvikles vil prosjektgruppa undersøke og vurdere om dette skal implementeres. Det skal også kartlegges og vurderes sammen med potensielle prosjektdeltakere om det er mulig å gjennomføre off-line/on-line målinger og om det er generatorer som skal rehabiliteres eller skrotes - og som kan inngå som prøveobjekter i prosjektet.

To partnere fra HydroStator ønsket å være med som finansierende partner inn mot en KSP-søknad, men de to andre finansierende partnerne fra HydroStator trakk seg eller unnlot å respondere på forslaget. Det ble derfor ikke sendt inn noen KSP-søknad i 2022.

Prosjektideen er beskrevet i en brosjyre (Vedlegg 1).

Lastsykling

En multitress testrigg ble bygget for å undersøke effekten av mange start-stopp (lastsykling) på høyspenningsisolasjon i vannkraftgeneratorer. For å etterligne virkelige start-stopp ble statorstaver utsatt for akselererte temperatursykluser ved å påtrykke 3,5 kA strøm og deretter kjøle ned med vifter mens høyspenning ble påtrykt samtidig. Elektriske delutladninger (*partial discharge, PD)* ble målt online under lastsykling. Etter hver 50. syklus ble off-line dielektriske tap og PD-målinger utført. Det var en tilsynelatende økning i dielektriske tap etter at stavene hadde blitt utsatt for totalt 250 sykluser. Under lastsykling avtok PD-nivået generelt med temperaturen, men det var en midlertidig økning i PD-aktivitet under temperaturstigning og -fall. Denne trenden kan forklares med ulik termisk ekspansjon i kobber og isolasjonsmateriale (epoxy-mica), noe som tyder på at statorstavene som belastes dynamisk vil bli utsatt for betydelig høyere PD-nivåer enn en statorstaver under en jevn belastning og temperatur, noe som fører til ytterligere forringelse av ytelsen til isolasjonen.

Arbeidet benyttet data generert i HydroStator og er publisert i [1]. Artikkelen er lagt ved som Vedlegg 2.

Mønstergjenkjenning

PD i høyspenningsisolasjonssystemer er både et symptom og årsak til forestående og endelige feil. Bruk av datadrevne metoder basert på PD-målinger vil gjøre det mulig for prediktive strategier å erstatte tradisjonelle vedlikeholdsstrategier. Dette arbeidet benytter maskinlæringsbaserte klassifiseringsmodeller for å identifisere og karakterisere PD-signaler som stammer fra laboratoriefremstilte, kunstige defekter i epoksy-mica materialeprøver. Tre forskjellige PD-kilder ble studert: overflateutladninger i luft, koronautladninger og utladninger forårsaket av indre hulrom/delamineringer. For å generere datasett av høy kvalitet for opplæring, validering og testing av klassifiseringsmodeller, ble faseoppløste PD-histogrammer (*phase resolved PD*, PRPD) for hvert testobjekt generert. Relevante statistiske og deterministiske egenskaper ble trukket ut for hver observasjon og ble merket basert på defekttypen (overvåket læring). Til slutt ble de trente og validerte ML-modellene brukt til å identifisere PD-kilder i statorviklingsisolasjonen tatt ut drift etter 50 år. *Support vector machine* (SVM), *ensemble* og *k-nearest neighbor* (kNN) algoritmer oppnådde signifikant høy nøyaktighet (≥ 95 %) av defektidentifikasjon.

Arbeidet ble gjort i samarbeid med HydroCen-prosjektet TwinLab, og benyttet også data generert i HydroStator. Arbeidet er publisert i [2] og lagt ved som Vedlegg 3.

Relevans og nytteverdi

Hensikten med KSP-forslaget HighFlex er:

- Forbedret tilstandsprediksjon og kunnskap om nedbrytning under dynamisk drift. Prosjektet er ment å gi tjenesteleverandører, generatoreiere og andre verktøy for bedre beslutningsgrunnlag for å utføre effektivt vedlikehold og kapitalforvaltning. Det vil være mulig å forlenge levetiden til generatorer eller planlegge for optimal rehabilitering eller fornyelse.
- Gi bedre kostnadsestimater for å gi fleksibilitet til nettet.
- Reduser antall fysiske inspeksjoner og reparasjoner, og reduserer dermed sannsynligheten for ulykker.
- Kunnskap om degradering under dynamisk drift vil bidra til andre interessenter, inkludert vindkraft, produksjonsindustri og elektrifisering av transport (skip, vei, luftfart).

Oppsummering

KSP-forslaget HighFlex ble utarbeidet, med gode tilbakemeldinger fra to av kraftselskapene som deltok i HydroStator. Det lyktes dessverre ikke å danne et tilstrekkelig stort konsortium og finansering til KSP-søknad i 2022. Prosjektforslaget er enda mer aktuelt nå ett år senere, siden fleksibiliteten til vannkraft er viktig for forsyningssikkerheten, og mer dynamisk drift av vannkraftanlegg er noe det planlegges for.

Relevante resultater på både dynamisk drift av vannkraftgeneratorer og datadrevne tilstandskontrollmetoder har blitt publisert gjennom dette forprosjektet, og vil i tillegg til å gjøre resultatene tilgjengelig for allmennheten også styrke en fremtidig KSP-søknad.

Referanser

[1] E. Eberg, E. Kantar, and T. G. Aakre, 'Multi-stress Cyclic Testing of Roebel Stator Bars for Hydropower', *Proceedings of the Nordic Insulation Symposium*, vol. 27, no. 1, Art. no. 1, Jul. 2022, doi: 10.5324/nordis.v27i1.4571.

[2] E. Kantar, J. M. Cascallo, T. G. Aakre, N. M. Thomsen, and E. Eberg, 'Use of Data-Driven Approaches for Defect Classification in Stator Winding Insulation', *Proceedings of the Nordic Insulation Symposium*, vol. 27, no. 1, Art. no. 1, Jul. 2022, doi: 10.5324/nordis.v27i1.4579.

Vedlegg 1

Multi-stress cyclic testing of Roebel Stator Bars for Hydropower

Espen Eberg[1], Emre Kantar[1] and Torstein Grav Aakre[1,2]

[1] *Department of Electric Power Technology, SINTEF Energy Research*

[2] *Department of Electric Power Engineering, Norwegian University of Science and Technology*

Abstract

A multi-stress test rig was built to investigate the effect of many start-stops (load cycling) on hydropower generator insulation. To emulate real-life start-stops, stator bars were subjected to accelerated temperature cycling by circulating a 3.5 kA current and then cooling down with fans while high voltage (service voltage) was applied simultaneously to the insulation. A total of 250 load cycles were applied. Partial discharges (PD) were recorded on-line during load cycling, and after every 50 cycles, off-line dielectric loss and PD measurements were conducted. There was an apparent increase in dielectric losses after they had been subjected to a total of 250 load cycles. During load cycling, the PD level generally decreased with temperature, but there was a temporary increase in the PD activity during temperature rise and fall. This trend can be explained by different thermal expansion in copper and insulating material (epoxy-mica), suggesting that a stator bar that is load-cycled will be exposed to substantially higher PD levels than a stator bar under a uniform load and temperature, leading to further deterioration of the insulation quality.

1. Introduction

Increased penetration of wind and solar power combined with phasing out of nuclear and carbon-based power plants will promote more intermittent hydropower operation to ensure the power grid's stability and maximize earnings. More frequent start/stops will introduce increased thermo-mechanical stresses on the components, especially in materials with different thermal expansion coefficient interfaces. Degradation and aging are thus expected to increase significantly for the stator windings, where copper bars or coils are wrapped with thermoset insulation, i.e., mica-polyester or mica-epoxy. During temperature increase and decrease, the copper in the bar changes its dimension to a larger extent than the insulation leading to thermo-mechanical stresses in the insulation material.

According to IEC 60034-18-34, [1] thermal cycling from the intermittent operation of generators can induce the following defects in the insulation layer:

1. Delamination between layers of insulation
2. Delamination between insulation and conductor
3. Abrasion of the outer surface (outer corona protection, OCP) of the insulation
4. Circumferential cracking of the insulation outside the slot part
5. Mechanical damage due to distortion of the end turns of winding

New generator windings must undergo a load cycling test procedure followed by a long-term high-voltage endurance test where the performance is compared to equivalent non-cycled stator windings to be qualified according to IEC 60034-18-34 [1], and IEEE Std 1310 [2] in North America. Diagnostic tests such as measurement of $\tan(\delta)$ and partial discharge (PD) measurements should also be undertaken, but there are no pass/fail criteria. These tests are proposed to facilitate identifying the dominating process leading to the degradation or failure of the stator windings.

Both $\tan(\delta)$ and PD magnitude have been measured periodically during several laboratory load cycling experiments to estimate the degradation and remaining lifetime of stator windings [3–5]. It has been reported that $\tan(\delta)$ increases with the number of load cycles due to increased thermo-mechanical stress giving delaminations between conductor and groundwall insulation [3]. Another study shows that $\tan(\delta)$ values also increase with the number of load cycles and correspond to a decreased breakdown voltage [4]. In [5], delaminations were observed by tapping on the insulation surface after 300 load cycles, and the PD activity increased correspondingly.

Multi-stress testing, with simultaneous thermal and electrical stress, on stator windings has been performed in [6]. They found that PD magnitude and current increased with the number of load cycles, caused by both increased size and number of delaminations and voids. But, after 1500 load cycles at maximum temperature 122 °C, there was only a small reduction in breakdown voltage. These studies indicate that new stator windings can endure several thousand load cycles.

A recent study [7] of stator bars aged under service with 10,000 start-stops investigated the effect of laboratory load cycling according to IEEE Std 1310, equivalent to IEC 60034-18-34. No delamination was observed before load cycling. After only 20-30 load cycles in laboratory severe delamination was found in companion with a significant increase in $\tan(\delta)$.

The average age of Norwegian hydro generators is approximately 50 years [8]. Even if few failures

occur annually, international statistics point to insulation system failures as the root cause [9]. The resilience to harsher and more dynamic loading patterns for these hydro generators, which have been in service for several decades, is unknown. No studies, to our knowledge, have investigated how multi-stress load cycling affects the condition of the insulation in generator bars taken from service, as they have either have been conducted on new stator bars or only have applied thermal cycling. There is thus a need to develop methods that can estimate the degradation and technical state of the insulation system providing real-life stress on the insulation system.

A multi-stress test rig was built to investigate the effect of many start-stops, i.e., load cycling, on the degradation of the stator winding insulation. To verify and demonstrate the rig's abilities and get reference measurements, new Roebel stator bars with wet bands (resin rich) mica-epoxy mainwall insulation and a nominal phase voltage (U_0) of 7.4 kV were tested in this work. The focus was on two main tasks for experiments with load cycling: i) A better understanding of how stresses/forces degrade the insulation during simultaneous load cycling and electrical stress, and ii) how do these stresses affect the insulation condition after many load cycles?

2. Equipment and methodology

2.1. Load cycling rig

In Fig. 1 (a) a sketch of the load cycling rig is shown, whereas Fig. 1 (b) shows an image of the physical implementation. We constructed the rig in a manner to be as flexible as possible so that it can accommodate different types/sizes of stator bars. The model stator was thus constructed of steel plates clamped with bolts around each stator bar. The stator bars were connected in series and formed a current loop at ground potential. AC current was provided by a split-core transformer, as seen in the front in Fig 1 (b). The model stator steel plates, which were isolated from the steel frame, were energized with high voltage (HV) from a medium voltage transformer, seen to the right in Fig 1 (b). This approach enabled temperature sensors to be inserted in the conductor (at ground potential) to provide accurate temperature measurements for regulation and control. Industrial fans were installed in the supporting steel frame and provided cooling. The test rig was controlled by a PLC system, and current and temperatures were logged at the center of the bar, at the end of the straight part, and in the end windings in both conductor and surface of the insulation.

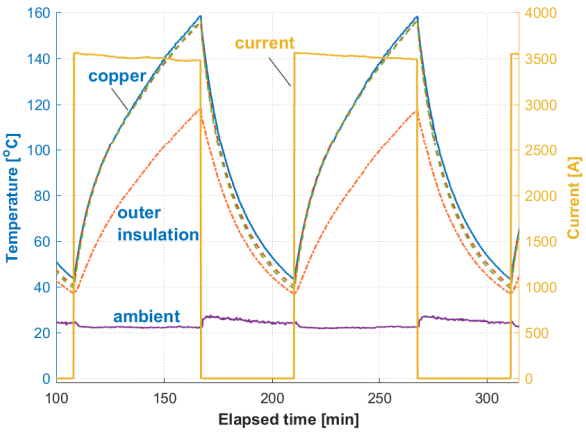
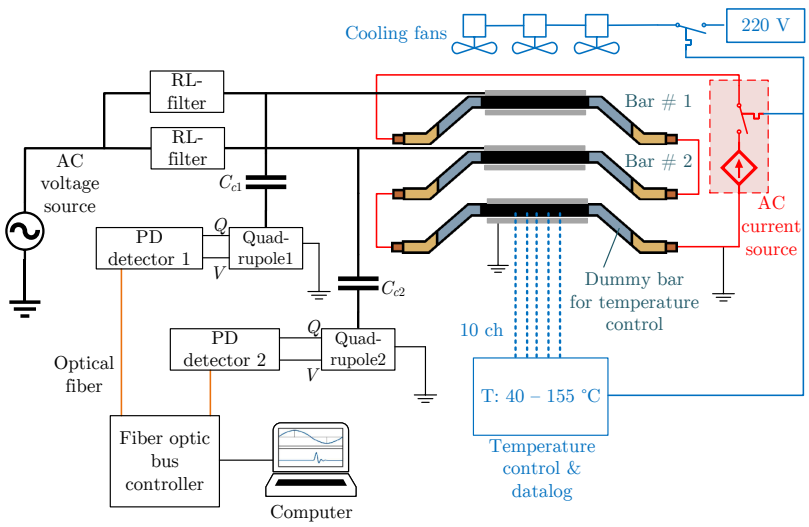
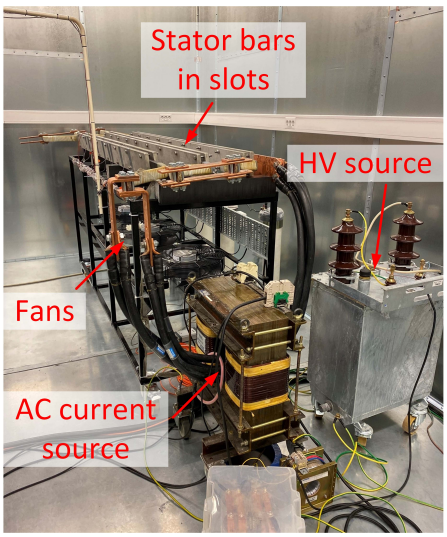


Fig. 2 – Load and temperature in the conductor, stator bar surface, and ambient during load cycling.

In order to emulate real-life start-stops, stator bars were subjected to accelerated temperature cycling by inducing current up to 3500 A and then cooling down with fans and HV simultaneously. A complete temperature cycle from 40−155 °C and back to 40°C took approximately 90-100 minutes, in accordance with IEEE Std 1310. This test protocol is significantly more stringent than



(a) Illustration of the constructed load cycling test set-up.



(b) Image of the constructed load cycling test set-up.

Fig. 1 – Multi-stress load cycling rig built for on-line and off-line tests.

a usual start/stop in a Norwegian hydropower plant, where the maximum temperature usually is $80-90\,^{\circ}\mathrm{C}$. To emulate the stator core, stainless steel plates were clamped on the sides of the stator bars. Three stator bars were simultaneously heated and cooled. Two bars were simultaneously electrically stressed under HV, while the third bar was at ground potential, allowing direct temperature measurements as a reference. The measured conductor temperature was used to regulate the auto start-stops in the test rig.

A total of 250 cycles (1 cycle = heating + cooling) were applied. During the load cycling, partial discharge (PD) in each bar was recorded on-line on individual channels. After every 50 cycles, the insulation conditions of the bars were characterized using dielectric loss and PD measurements (off-line). The testing scheme is shown in Fig. 3.

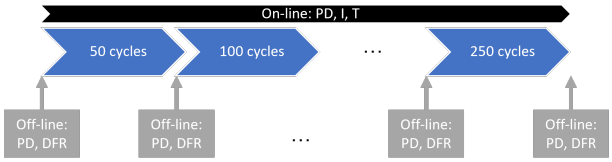


Fig. 3 – Outline of the testing scheme.

2.1.1. Partial discharge measurements

PD measurements were performed according to IEC 60270 [10], using an Omicron MPD 600 system and 800 pF coupling capacitors. During load cycling, the two bars energized with HV were electrically isolated by RL-filters in the high frequency (HF) band, as shown in Fig. 1(a), thus allowing for separate PD measurements of each stator bar (without crosstalk) during load cycling. At intervals of 50 load cycles, the end connections were removed, and each stator bar was measured off-line to also investigate inception voltage (PDIV) and voltage dependency of PD-parameters.

2.1.2. Dielectric loss measurements

Dielectric frequency response (DFR) was measured at intervals of 50 load cycles with a GE-Programma IDA 206 Insulation Diagnostic Analyzer connected to a Trek HV amplifier. RMS test voltages were 140 V and from 0.2 U_0 to 1.2 U_0. Since the stator bars were situated in the test rig during the measurements, placing a guard at the end-corona protection (ECP) was not possible. Hence the losses in the ECP were included in the measurements, which can cause a significant increase in measured losses [11]. Only relative changes should thus be studied rather than the absolute values from the measurements with different guarding arrangements.

2.1.3. Test objects

The test objects in this study were pristine Roebel bars with 2.3 m slot length and rated at $U_0 = 7.4$ kV (phase-to-ground). The mainwall insulation system is resin-rich epoxy-mica, and ECP was painted following the supplier's guidelines. Three stator bars were load-cycled. One bar served as a reference for temperature measurements and was not at HV potential, while two bars were at HV potential simultaneously denoted G009 and G010 in this study.

3. Results and discussion

3.1. On-line PD

In Fig. 4 the PD-magnitude and PD repetition rate, n, for stator bar G009 and conductor temperature for the reference bare during two complete load cycles are shown. It can be observed that both magnitude and intensity are largest during increasing or decreasing temperature, with peak magnitudes of 6.5 nC and 7.3 nC, respectively. The PD magnitude stabilizes at 50 pC above $140\,^{\circ}\mathrm{C}$, while there is no plateau at low temperatures. Since the PD magnitude is much larger during temperature change than at steady-state levels at room and elevated temperatures (not shown here), the observed effect was likely caused by the formation of voids or changes in the geometry of existing voids within the insulation. As the stator bar heats up, there is an elongation that is not reflected in the thermoset insulation, and voids increase in size or the formation of new voids occurs. Likely competing mechanisms at high temperatures are the increased conductivity and pressure increase, which promote a decrease in PD magnitude. At a certain temperature, one or both of these mechanisms will cause a reduction in PD magnitude, approximately at $100\,^{\circ}\mathrm{C}$ in the current experiment. As the temperature drops, the pressure in the voids and conductivity in the insulation and void wall decrease, and thus the PD activity increases again.

The increased PD activity during large temperature changes manifests that load cycling without simultaneously applied HV will not represent the realistic stresses the mainwall insulation undergoes during start/stops in a hydropower station. Since only the thermo-mechanical effects are engaged, and this can lead to an over-optimistic conclusion regarding the lifetime of generator windings.

3.2. Off-line PD

The load cycling was stopped after each 50 load cycle to perform off-line low-noise PD measurements to evaluate if there had been any subtle changes or degradation to the insulation system during the load cycling that was not observed by the on-line assessment. In Fig. 5, the total apparent charge per period, 99 percentile maximum charge, repetition rate, and average discharge are shown for each 50 load cycle.

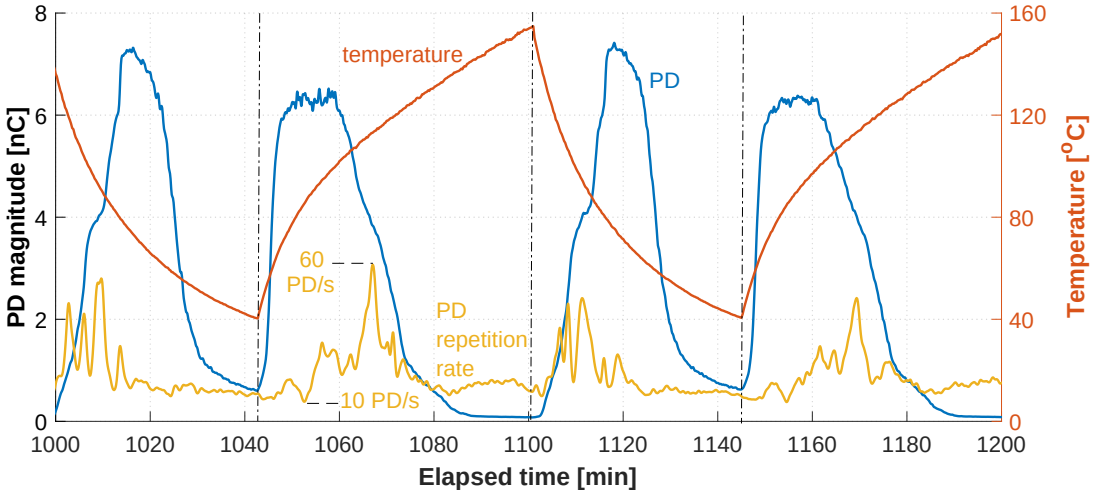


Fig. 4 – PD magnitude and intensity during load cycling.

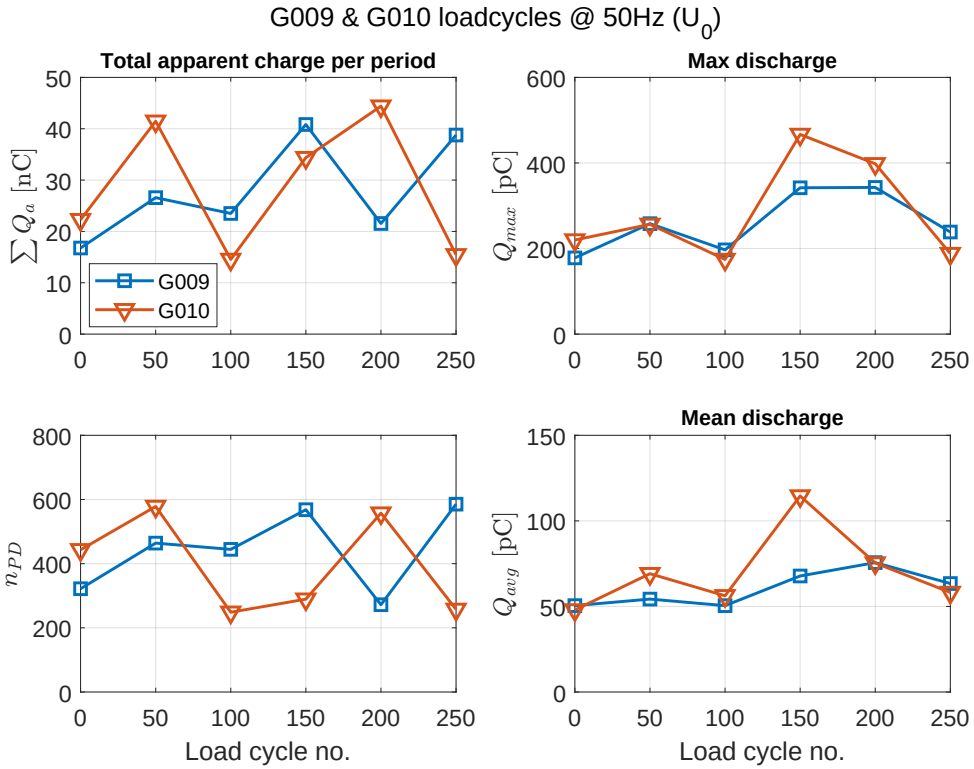


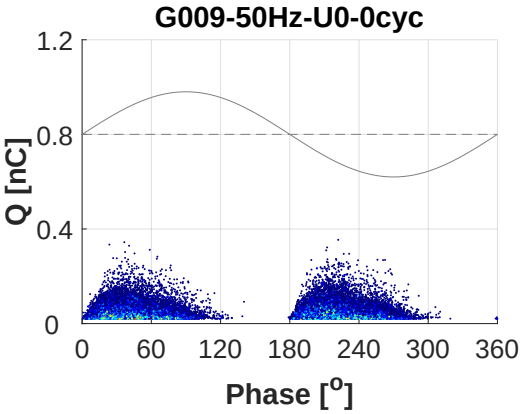
Fig. 5 – Trend of PD parameters after 0, 50, ..., 250 load cycles.

The only observable change in the trend is an increase in the max and average discharge at 150 load cycles. This was caused by a circumferential crack around the end-winding, as shown in Fig. 6. Due to the curvature in the end-winding, there will be variations in mica tape overlap and bonding between layers, which can cause decreased mechanical strength of the mainwall insulation in this region. Combined with the elongation of the copper core, this type of circumferential cracking can be the result. As load cycling proceeded, the increase in PD-activity declined back to normal. A similar trend was observed in the on-line measurements.

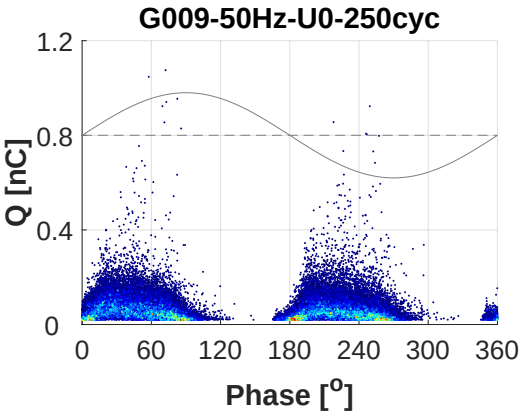
In Fig.,7 the phase-resolved PD (PRPD) histograms before load cycling (a) and after 250 load cycles (b) are

Fig. 6 – Crack in end-winding.

shown for stator bar G009. In these histograms, all PD events are shown and constitute the source data for the trend of PD parameters shown in Fig. 5. After 250 load cycles, it can be seen that there are PD events with a larger magnitude compared to before load cycling. This is consistent with an increase in void size [12]. It is also seen that after 250 load cycles, there are discharges present at the trailing edge just prior to polarity reversal in the PRPD pattern, which is caused by an increase in remnant charges in the voids. The latter can be a change in the electronic states at the surface or merely because larger deposited charges need a longer time to decay to zero. Combined, these observations suggest that there has been a change or degradation in the insulation system of the generator bars. They are not observed in the trend plot because the largest PD magnitudes are very few compared to the bulk of small PDs.



(a) PRPD histogram prior to load cycling.



(b) PRPD histogram after 250 load cycles.

Fig. 7 – PRPDA histograms before and after load cycling.

3.3. Dielectric frequency response

For both stator bars, there was a clear increase in dielectric losses measured at U_0 after they had been subjected to the 250 load cycles. In contrast, no consistent increase could be seen for the dielectric loss measured at low voltage (140 V), as shown in Fig. 8. Below PDIV, an increase in $\tan(\delta)$ will generally probe the overall increase in conductivity from the degradation of the insulation system. In contrast, above PDIV, there will also be a contribution to losses from the PD current, and an increase in $\tan(\delta)$ is expected. PDIV for epoxy-mica stator bars is usually in the range of $0.6U_0 - 0.8U_0$, see e.g. [13]. This is consistent with increased void size, i.e., more delamination of the mainwall insulation, as reported in the literature [3], and also with the increased PD-magnitude after 250 load cycles as shown in Fig. 7. Degradation of the ECP could also be a source of increased losses, but since it was not possible to guard the ECP in the measurements, the contribution from the ECP could not be isolated.

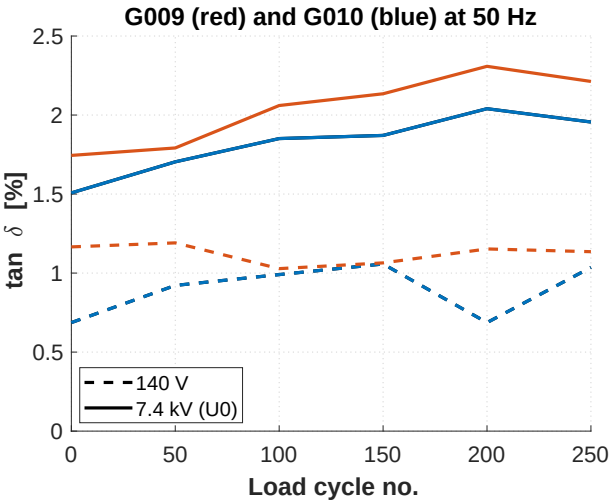


Fig. 8 – $\tan(\delta)$ as function of number of applied load cycles, measured at 140 V (dotted line) and U_0 (solid line).

4. Summary and outlook

A novel multi-stress test rig, with simultaneous thermal and electrical stress, for generator stator bars, has been constructed and demonstrated by completing 250 load cycles with simultaneous application of HV potential on two stator bars.

There is a significant magnitude of PD activity during temperature increase and decrease, showing that there potentially is additional electric-stress induced degradation during load cycling that current IEEE and IEC standards do not probe. Trending PD parameters show little change during load cycling, but careful inspection of PRPD histograms reveal that there are more, but few, PDs with large magnitude after 250 load cycles. This is consistent with the increase in $\tan(\delta)$ measured at U_0 present after 250 load cycles.

Following IEEE and IEC standards, load cycling is done as part of the test and development regime for new stator bars. The test is accompanied by monitoring changes in $\tan(\delta)$, and the acceptance criteria is that no significant change in $\tan(\delta)$ or $\Delta \tan \delta$ should occur – the latter being the difference in $\tan(\delta)$ between to following voltage levels. It is shown in this work that there is, in fact, an increase after 250 load cycles for new, previously unused stator bars. Following IEEE and IEC standards, load cycling tests do not include simultaneous application of HV to the mainwall insulation. Combined

with the fact the large magnitude was observed during temperature change, we suspect that the more realistic multi-stress testing is the main cause, and such testing, therefore, should be considered to impose more realistic testing regimes.

The test rig and methods for assessment have now been demonstrated. Future work should focus on accelerated multi-stress ageing of stator bars taken from service in order to investigate how old hydro generators will cope with more intermittent operation. Comparison of load cycling with and without simultaneous HV should be of interest for standardization bodies, as this will address the relevance of the existing standards.

Acknowledgment

This work is funded by the project "Hydrogenerator Stator Winding Insulation Assessment." The project is supported by The Research Council of Norway (Project No. 255099/E20) and industrial partners.

References

[1] IEC 60034-18-34:2012 - Functional evaluation of insulation systems - Test procedures for form-wound windings - Evaluation of thermomechanical endurance of insulation systems. Technical report, International Electrotechnical Commision, 2012.

[2] IEEE std 1310-2012 - IEEE Recommended Practice for Thermal Cycle Testing of Form-Wound Stator Bars and Coils for Large Rotating Machines. Technical report, May 2012.

[3] M. Farahani, H. Borsi, E. Gockenbach, and M. Kaufhold. Partial discharge and dissipation factor behavior of model insulating systems for high voltage rotating machines under different stresses. *IEEE Electrical Insulation Magazine*, 21(5):5–19, September 2005.

[4] H. Mitsui, K. Yoshida, Y. Inoue, and S. Kenjo. Thermal Cyclic Degradation of Coil Insulation for Rotating Machines. *IEEE Transactions on Power Apparatus and Systems*, PAS-102(1):67–73, January 1983. Conference Name: IEEE Transactions on Power Apparatus and Systems.

[5] H. Zhu, C. Morton, and S. Cherukupalli. Quality evaluation of stator coils and bars under thermal cycling stress. In *Conference Record of the 2006 IEEE International Symposium on Electrical Insulation*, pages 384–387, June 2006. ISSN: 1089-084X.

[6] R. Bartnikas and R. Morin. Multi-stress aging of stator bars with electrical, thermal, and mechanical stresses as simultaneous acceleration factors. *IEEE Transactions on Energy Conversion*, 19(4):702–714, December 2004.

[7] Marcelo Jacob da Silva, Fernando Pereira, Joao Fiorini Cordeiro, Tomas Erikson Lamas, Camila Goncalves dos Santos, Tamy Koreeda Aoki, Egidio Jose Faria, and Thomas Hildinger. Facts and Artifacts from IEEE 1310-2012. In *2018 IEEE Electrical Insulation Conference (EIC)*, pages 170–174, San Antonio, TX, June 2018. IEEE.

[8] T. M. Sneve. Aldersfordeling for komponenter i kraftsystemet, levetid og behov for reinvesteringer. Technical Report NVE 8, 2005.

[9] Survey on Hydro Generator Instrumentation and Monitoring. Technical Report TB 682, CIGRE WG A1.40, 2009.

[10] IEC 60270:2000+AMD1:2015 CSV - High-voltage test techniques - Partial discharge measurements. Technical report, International Electrotechnical Commision, 2015.

[11] M.G. Krieg-Wezelenburg. Dielectric Dissipation factor measurements on stator insulation - results from a global survey. In *2020 IEEE Electrical Insulation Conference (EIC)*, pages 269–273, June 2020.

[12] F. Gutfleisch and L. Niemeyer. Measurement and simulation of PD in epoxy voids. *IEEE Transactions on Dielectrics and Electrical Insulation*, 2(5):729–743, October 1995.

[13] E. Eberg, T. G. Aakre, G. Berg, and S. Hvidsten. Comparison of Offline VLF PD Measurements and Online PD Measurements on a 50-Year-Old Hydrogenerator Stator in Norway. In *2018 IEEE Electrical Insulation Conference (EIC)*, pages 542–546, June 2018.

Use of Data-Driven Approaches for Defect Classification in Stator Winding Insulation

Emre Kantar[1], Jaume M. Cascallo[2], Torstein Grav Aakre[1], Nina M. Thomsen[1], and Espen Eberg[1]
[1] *Department of Electric Power Technology, SINTEF Energy Research*
[2] *Department of Electric Power Engineering, Norwegian University of Science and Technology*

Abstract
Partial discharges (PD) in the high voltage insulation systems are both a symptom and cause of terminal and impending failures. The use of data-driven methods based on PD measurements will enable predictive strategies to replace traditional maintenance strategies. This paper employs machine learning-based classification models to identify and characterize PD signals originating from lab-made artificial defects in epoxy-mica material samples. Three different PD sources were studied: surface discharges in air, corona discharges, and discharges caused by internal cavities/delaminations. To generate high-quality datasets for the training, validation, and testing of classification models, Phase-Resolved PD (PRPD) data for each test object was obtained at room temperature under 50 Hz AC excitation at 10 % above the PD inception voltage (PDIV) of each sample. Relevant statistical and deterministic features were extracted for each observation and were labeled based on the defect type (supervised learning). Finally, the trained and validated machine learning (ML) models were used to identify PD sources in the service-aged stator winding insulation. Support vector machines (SVM), ensemble, and k-nearest neighbor (kNN) algorithms achieved significantly high accuracy ($\geq$ 95 %) of defect identification.

1. Introduction
Intermittent power generation causes stator winding insulation of hydrogenerators, which were designed for primarily continuous operation 50 years back, to experience damaging and frequent service failures, resulting in long downtimes, and costly repairs, thus, significant economic losses [1, 2]. Reliable and accurate condition monitoring of stator winding insulation has been widely performed using both off-line and on-line partial discharge (PD) measurements. PD measurements are a vital tool for assessing and monitoring the condition of power equipment. Different sources of PD have different effects on the insulation performance and thus reliability of the power apparatus. Therefore, identifying various PD sources at different locations is of great importance for the health assessment of stator winding insulation [3].

Data-driven methods employing artificial intelligence, such as machine learning (ML) algorithms, are now more feasible due to increasing computation power and accessible tools, and a growing interest in such quantitative and predictive methods is expected. Developing such models for generator lifetime/condition qualification and estimation will accelerate the transition from traditional strategies toward predictive strategies, such as upgrading insulation components before they are estimated to fail [4].

Human experts usually perform a condition assessment of generator insulation with experience and qualitative judgment of data. The identification of PD sources (and their severity) is usually done using Phase-Resolved PD (PRPD) analysis, where each PD event is resolved into the apparent discharge magnitude (Q_a), phase angle (ϕ), and the number of the PDs (n). The use of ML-based data-driven models as a decision support tool can maximize the reliability and accuracy of the defect identification by unlocking hidden correlations.

ML-based PD diagnostics have been receiving increasing attention in the literature to handle the growing amount of data, reduce the human labor for feature engineering and tap the full potential of the data [5]. However, most of these studies employ similar fundamental features and are limited to specific test setups, whose experimental conditions do not explicitly state systematical details. These limitations inhibit comparative analysis and reproducibility. To further develop robust ML-based models, different statistical and deterministic PD features and due details for data acquisition and feature extraction techniques should be introduced.

The primary purpose of this study is to employ novel features (predictors) extracted from the obtained PRPD datasets and generate high-quality datasets for the training, validation, and testing of classification models based on various ML algorithms (classifiers). To that end, lab-made artificial defects are made to represent the most common defect types. Then, the qualified classification models (trained and tested) are used to predict possible defect types in the service-aged stator winding insulation.

2. Methodology
2.1. Test Objects
Test samples with known defects were made to emulate the most common discharge sources encountered in practice, which are classified into three main groups:

i. corona discharges in the air (e.g., semiconductor-field grading paint (or tape) junction at the end windings),

ii. surface discharges (e.g., slot discharges),
iii. internal discharges (e.g., internal cavities and delaminations).

Fig. 1(a) illustrates the arrangement to induce corona discharges in an air gap (4 cm) between an energized piece of thin aluminum wire (with a tip diameter of 1 mm) and a flat ground electrode. To induce surface discharges, a similar thin wire was attached to the electrically grounded semiconductive coating and was extended towards the high voltage (HV) copper strands of a real hydrogenerator stator bar (≈ 2 cm gap), as depicted in Fig. 1(b). In this work, stator bar is used interchangeably with stator winding insulation. Also, three service-aged stator bars (similar to the one shown in Fig. 1(b) without the wire) were used to predict PD sources associated with them.

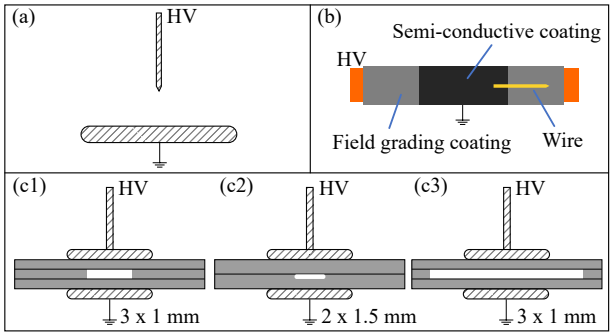


Fig. 1 – Illustration of the lab-made samples as artificial PD sources. (a) Corona discharges (rod to plane). (b) Surface discharges. (c) Void discharges: c1. Center void: 10 mm void diameter and 1 mm void gap distance, c2. Delamination closer to the bottom electrode, and c3. Large void delamination near end-windings (voids/delaminations with unstressed walls).

We made laboratory objects from resin-rich mica/epoxy/glass-fiber tape to emulate internal discharges originating from voids and delaminations. The object dimensions were 100 mm $\times$ 100 mm $\times$ 3 mm. 1-mm of sheet/plate thickness was formed by stacking six tape layers half-overlapped. We cured the test objects at 160 °C for one hour under pressure. Metal spacers were used to create cylindrical voids during the curing process. Upon pressing cured plates together, a test object with a definite void dimension was formed.

Three different test objects with a total insulation thickness of 3 mm incorporating different void types were made, as exhibited in Fig. 1: (c1) 10 mm void diameter and 1 mm void gap distance, (c2) 5 mm void diameter and 0.5 mm void gap distance with round edges, and (c3) 40 mm void diameter and 1 mm void gap distance. The main difference between the voids in (c1) and (c3) is that the void in (c3) has electrically unstressed walls and represents large delaminations and voids. In (c2), the void was created only on one plate while the other plate was void-free, i.e., the void is slightly closer to the bottom electrode, representing an asymmetrical void in the insulation. The diameter of the HV electrode was 30 mm.

2.2. Test Setup

We generated high-quality PD datasets (free of crosstalk and noise above a selected threshold value of 1 pC) for the training, validation, and testing of various ML algorithms using a commercial PD acquisition unit (Omicron MPD 600) along with a 100-pF coupling capacitor. Fig. 2 illustrates the HV PD setup used for data generation, and *test object* stands for the lab-made samples with artificial defects. Desired voltage amplitude and frequency were set by a digital to analog converter (DAQ) and amplified to HV by an HV amplifier (TREK 20/20C–20 kV, 20 mA) in series with an RLC-low-pass filter with a cut-off frequency at 5 kHz. Measurements were performed according to IEC 60270, [6] where the amplitude spectrum was integrated with a center frequency of 250 kHz and a bandwidth of 300 kHz.

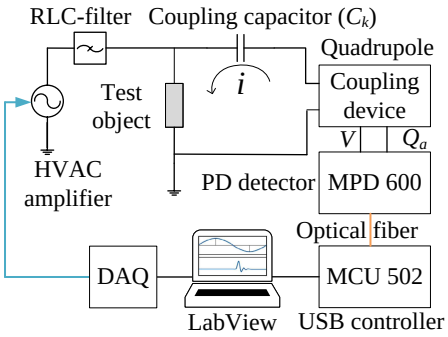


Fig. 2 – Illustration of the experimental PD measurement setup.

2.3. Test Procedure and Data Acquisition

The test objects were subjected to voltage preconditioning before the data collection activity to preempt the so-called "memory effect", and thus, the same test sample could be tested multiple times to achieve reliably extensive training and testing datasets. PD inception voltage (PDIV) of each test object had been experimentally determined in accordance with the definition in IEC 60270, and test objects were subjected to voltage preconditioning at 10 % above their corresponding PDIV at 50 Hz for 5 minutes before each test was performed (data collection activity). Given the stochastic nature of the PD phenomenon, generating data using the same test sample should be performed at least five times for acceptable statistical significance [7] as well as to generate an extensive and reliable dataset.

For this purpose, each sample was subjected to HV for 600 s at 50 Hz AC voltage. Each measurement (also referred to as "observation") incorporated 1000 AC cycles at 50 Hz (20 s), allowing for up to a total of 30 separate data windows (30 observations) for the same test sample (30 $\times$ 20 s $=$ 600 s). Fig. 3 presents the steps followed for data and feature extraction, training, validation (fraction of training data used during training for tuning hyperparameters), testing, and prediction. MATLAB's inherent functions and Classification Learner Toolbox were used for the entire data analysis and ML tasks. Critical remarks are listed below.

1. Employing the obtained PRPD data, statistical and PD quantities were extracted using 20 observations for each test sample out of 30 available observations: i.a., PD inception/extinction voltage, pulse repetition rate, phase angle, average discharge current, discharge power, quadratic rate, kurtosis, skewness. The generated parameters were labeled based on the defect type (class). The complete list of extracted features is shown in Table 1. The reasons for the selection of features are given in the next section.
2. The extracted features for each observation were labeled based on the artificial defect type (supervised learning). It should be highlighted that observations belonging to c1–c3 type defects were grouped/labeled together as "void discharge." The entire dataset was then split into training and testing sets using the hold-out method, where 80 % of the data was allocated for training, and the remaining 20 % of the data was used for testing. In addition, "k-fold cross-validation" was used on the training dataset where the dataset was arbitrarily divided into "k" groups. In this work, 5-fold cross-validation was used. One of the groups was used as the validation set, and the rest were used as the training set, as illustrated in Fig. 3. Then, the chosen model was trained on the training set and was evaluated on the test set. The procedure continued until each unique group was used as the test set.
3. Selected machine learning classifiers, such as support vector machines (SVM), k-nearest neighbor (kNN), ensemble, and decision tree, including their various subtypes, were trained by optimizing their hyperparameters. Then, predictions were made based on the defect type (label or class). Based on the training and test accuracies, the best models were chosen.
4. The qualified ML models were then used to predict/identify PD sources in service-aged (50 years) stator bars. Following the test procedure performed for the lab-made artificial defects, the same predictors/features were extracted from the obtained PRPD data of the stator winding insulation for each observation. The observations were then fed into the selected ML models, and possible PD sources were predicted based on the three training classes/labels.

The end-windings of these bars were removed due to the presence of asbestos, and then the semiconductive coating was removed at the ends, and field stress grading paint was applied by following the manufacturer's instructions. Thus, PD sources related to end-windings were eliminated. Also, the semi-conductive and field grading coating on the bars were checked against any abrasion or damage; only the undamaged bars were tested. Therefore, any PDs arising from the stator bars were expected to originate from internal discharges in the mainwall insulation represented by one or more void classes defined in c1–c3. However, the dataset was not labeled due to the uncertainty of unknown defects that might be present in the winding insulation. A detailed discussion of the predicted classes is given in the results.

2.4. Data Preprocessing and Feature Extraction

Firstly, the entire dataset was separately grouped for the positive and negative half-cycles as a reference to a sinusoidal AC voltage waveform to account for the asymmetrical PD events based on the defect type. The superscript $(+)$ for the positive half cycle and $(-)$ for the negative half cycle were used, respectively. Subsequently, numeric features (deterministic) such as number of charges, average, median, maximum (99th percentile employed to eliminate outliers) charge magnitudes, and discharge power were calculated for the positive and negative half-cycles (see Table 1) using the de-noised $\phi - Q_a - n$ data exported from MPD 600.

Secondly, statistical post-processing of the $\phi - Q_a - n$ data was performed in addition to the above-mentioned deterministic parameters. For the statistical analytics, PRPD data were characterized by using $\mathrm{H_n}(\phi)$–*phase distribution of the number of PDs* and $\mathrm{H_{qn}}(\phi)$–*phase distribution of mean discharge amplitude*. An example of such distributions extracted from the training database is shown in Fig. 4(a). Similarly, each feature was extracted separately for the positive and negative half-cycles using the corresponding $\mathrm{H_n}(\phi)$ and $\mathrm{H_{qn}}(\phi)$, respectively.

We employed Weibull distribution, kurtosis (Ku), and skewness (Sk) analysis for the statistical analytics. The distribution of the obtained discharge amplitudes (Q_a) follows an S-shape-like trajectory on the cumulative distribution function (cdf) plot, as presented in Fig. 4(b), indicating the existence of more than one active discharge mechanism [8]. The plot shows that a five-parameter mixed Weibull distribution (with two separate mechanisms) fitted the diffused discharge data well. A mixed Weibull distribution based on the sum rule is given by:

$$F(q) = p\,F_1(q) + (1-p)\;F_2(q)\,, \tag{1}$$

where q is the discharge magnitude (Q_a in our case), $F_1(q)$ and $F_2(q)$ are the cdfs of each discharge mechanism, p is the probability of occurrence of the sub-population $F_1(q)$ with $0 \leq p \leq 1$ [8]. Below are five parameters; α_1, β_1, α_2, β_2, and p, shown explicitly:

$$\begin{aligned} F(q) =& p\left\{1-\exp\left[-\left(\frac{q}{\alpha_1}\right)^{\beta_1}\right]\right\} + \\ & (1-p)\left\{1-\exp\left[-\left(\frac{q}{\alpha_2}\right)^{\beta_2}\right]\right\}, \end{aligned} \tag{2}$$

where α is the scale parameter, and β is the shape parameter of each cdf. It is assumed that $p = 0.5$.

The shapes of $\mathrm{H_n}(\phi)$ and $\mathrm{H_{qn}}(\phi)$ have characteristic features for different PD sources [9]. Skewness and kurtosis have widely been used for the quantitative classification of PDs in artificial defects [9, 10] and hydro generators [11]. In relation to a normal distribution, skewness refers to the degree of asymmetry of a distribution: Sk is positive (negative) for a leftward (rightward) shift of a normal distribution. On the other hand, kurtosis stands for the degree of the sharpness of

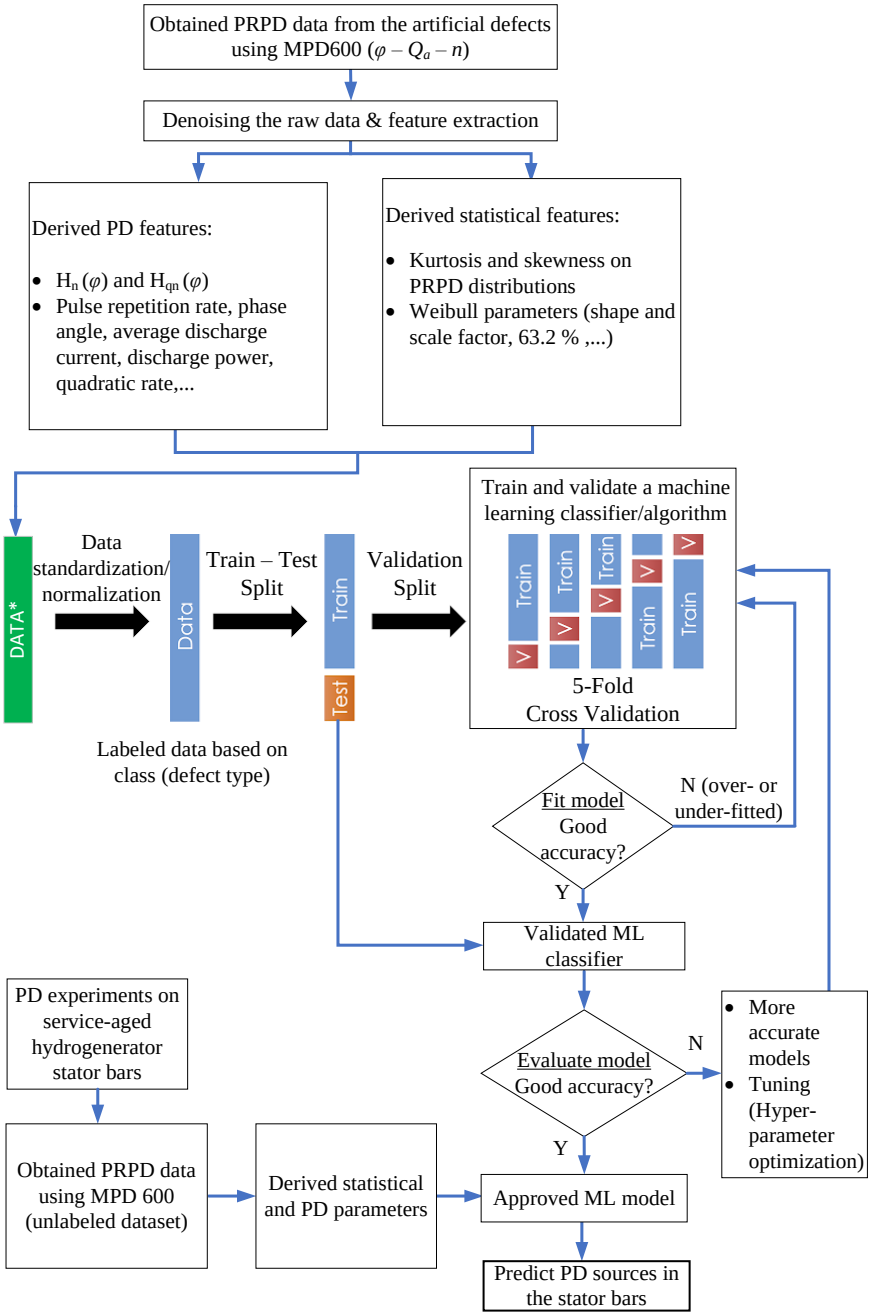


Fig. 3 – Flowchart showing the steps of data preparation, feature extraction, model training, testing, and tuning.

a distribution in relation to a normal distribution, and Ku is positive (negative) for a pointier (flatter) distribution compared to a normal distribution. Lastly, feature scaling was applied to the data to equalize their influence on the model. In this work, each feature column in the vector containing the entire observations (DATA* as delineated in Fig. 3) was scaled to limit its range in the interval $[0, 1]$.

3. Results and Discussion

The obtained PRPD plots for 1000 AC cycles (50 Hz) for each defect type and winding insulation are shown in Fig. 5. The algorithms were then trained and tested based on the extracted features from the PRPD data to distinguish different PD sources under AC voltage. The input data consisted of the standardized 32 features (Table 1). The predicted classes were rod to plane (*Rod2plane*), surface discharge (*surfaceDischarge*), and internal discharges (*Void*).

3.1. Cluster Analysis and Feature Selection

A quick way to check if the classes can be grouped based on their extracted features is to visualize the data on a scatter plot and see if there are any obvious patterns or groups. For this purpose, chosen features were plotted against each other on a 3D scatter plot, as shown in Fig. 6. The selected features in the plot are closely related to most of the other features; hence similar patterns were observed among the classes. The scatter plot suggests that the extracted features can be useful to differentiate between the PD sources because the

Table 1 – Features (predictors) extracted from the experimentally obtained PRPD datasets.

No.	Type	Feature (predictor)	Notation
1	Deterministic (PD)	Total number of pulses	nPulse
2	Deterministic (PD)	Number of positive charges	nPosCh
3	Deterministic (PD)	Average positive charge magnitude	avgPosCh
4	Deterministic (PD)	Median positive charge magnitude	medPosCh
5	Deterministic (PD)	Maximum positive charge magnitude	maxPosCh
6	Deterministic (PD)	Number of negative charges	nNegCh
7	Deterministic (PD)	Average negative charge magnitude	avgNegCh
8	Deterministic (PD)	Median negative charge magnitude	medNegCh
9	Deterministic (PD)	Maximum negative charge magnitude	maxNegCh
10	Deterministic (PD)	Number of pulses per cycle	ChRepRate
11	Deterministic (PD)	Average discharge current	avgCur
12	Deterministic (PD)	Total discharge magnitude per cycle	$\mathrm{sumQ_a}$
13	Deterministic (PD)	Average discharge power	avgP
14	Deterministic (PD)	Quadratic rate	D
15	Statistical (Weibull)	63.2% positive discharge magnitude	Qpos63
16	Statistical (Weibull)	63.2% negative discharge magnitude	Qneg63
17	Statistical (Weibull)	Scale parameter 1 $(+)$	ScalePos1
18	Statistical (Weibull)	Scale parameter 2 $(+)$	ScalePos2
19	Statistical (Weibull)	Shape parameter 1 $(+)$	ShapePos1
20	Statistical (Weibull)	Shape parameter 2 $(+)$	ShapePos2
21	Statistical (Weibull)	Scale parameter 1 $(-)$	ScaleNeg1
22	Statistical (Weibull)	Scale parameter 2 $(-)$	ScaleNeg2
23	Statistical (Weibull)	Shape parameter 1 $(-)$	ShapeNeg1
24	Statistical (Weibull)	Shape parameter 2 $(-)$	ShapeNeg2
25	Statistical (Kurtosis)	Phase distribution of number of discharges $(+)$	$\mathrm{Ku\ (H_n^+)}$
26	Statistical (Kurtosis)	Phase distribution of mean discharge amplitude $(+)$	$\mathrm{Ku\ (H_{qn}^+)}$
27	Statistical (Kurtosis)	Phase distribution of number of discharges $(-)$	$\mathrm{Ku\ (H_n^-)}$
28	Statistical (Kurtosis)	Phase distribution of mean discharge amplitude $(-)$	$\mathrm{Ku\ (H_{qn}^-)}$
29	Statistical (Skewness)	Phase distribution of number of discharges $(+)$	$\mathrm{Sk\ (H_n^+)}$
30	Statistical (Skewness)	Phase distribution of mean discharge amplitude $(+)$	$\mathrm{Sk\ (H_{qn}^+)}$
31	Statistical (Skewness)	Phase distribution of number of discharges $(-)$	$\mathrm{Sk\ (H_n^-)}$
32	Statistical (Skewness)	Phase distribution of mean discharge amplitude $(-)$	$\mathrm{Sk\ (H_{qn}^-)}$

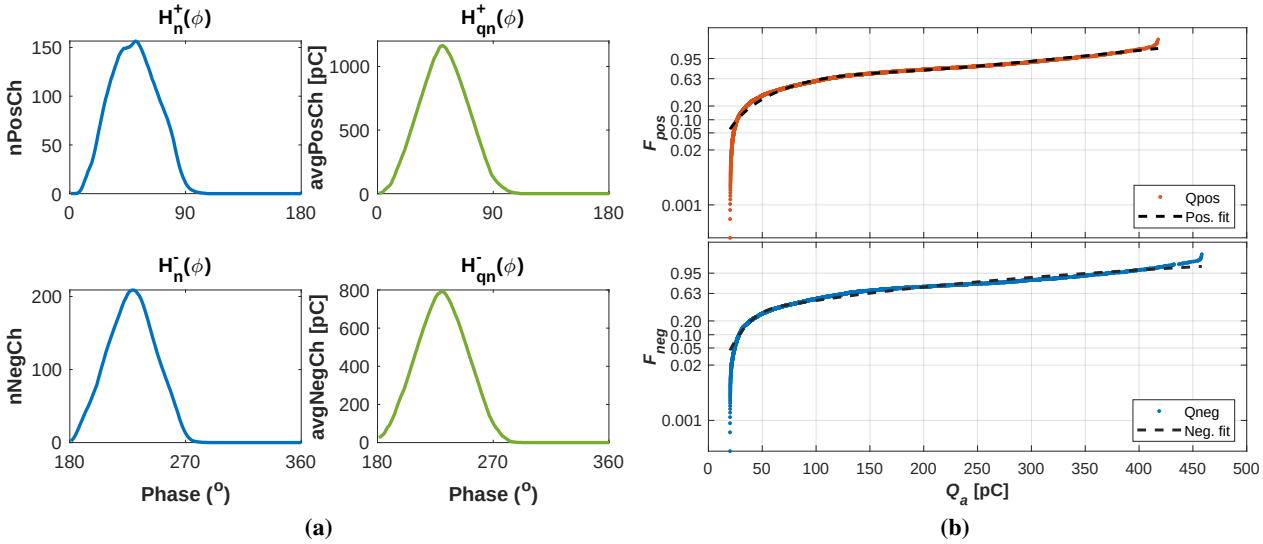


Fig. 4 – (a) Example of extracted $\mathrm{H_n}(\phi)$ and $\mathrm{H_{qn}}(\phi)$ distributions from the PRPD data of void discharge. (b) Fitted mixed Weibull cdf plots for positive and negative half-cycles retrieved from the PRPD data of void discharge.

data points belonging to each class form distinguishable clusters. Also, using a box plot to compare classes after a statistical Weibull parameter, viz. ShapeNeg2 (β_2 of the negative half cycle data) suggests discernible differences, as presented in Fig. 7.

3.2. Classification Analysis

The performances of the trained and tested algorithms are usually interpreted using a confusion matrix, which depicts the amount of correctly classified instances in the major diagonal and the wrong predictions in the minor diagonal. Also, "accuracy" is a useful parameter showing the overall performance and is defined as the value obtained by dividing the number of correct predictions by the number of total predictions.

Table 2 shows a list of selected classifiers that were trained and validated along with their accuracies. Optimized versions for each classifier stand for the performed hyperparameter optimization (model tuning),

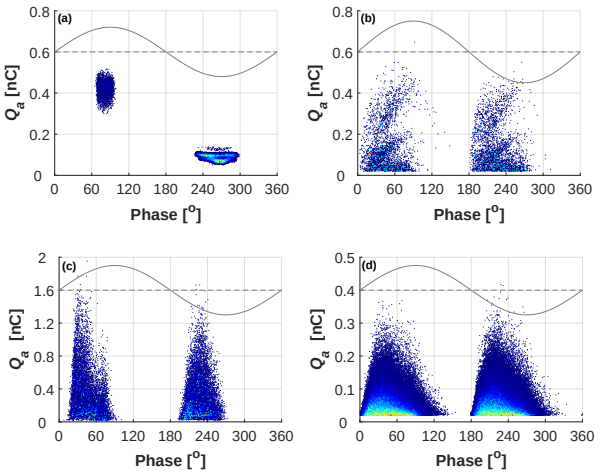


Fig. 5 – Obtained PRPD from the lab-made samples. (a) Rod to plane. (b) Surface discharge. (c) Void discharge. (d) Discharges in the stator winding insulation.

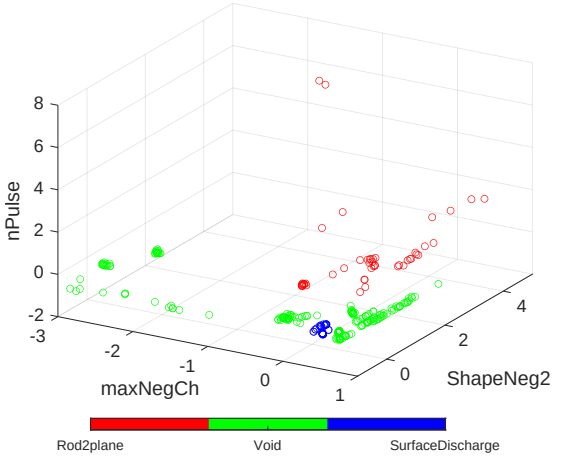


Fig. 6 – Scatter plot of selected features to visualize clusters.

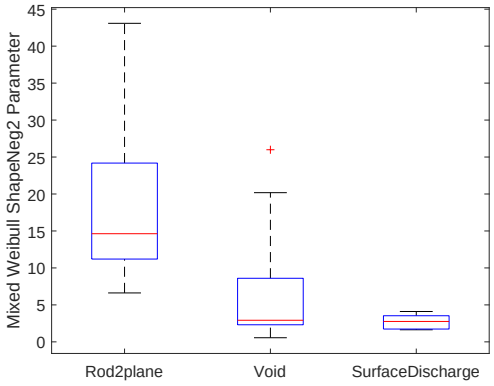


Fig. 7 – Box plot for the shape parameter (ShapeNeg2) vs. classes.

as illustrated in Fig. 3. As an example, Fig. 8 presents the confusion matrices for validation and test results of the tuned kNN-classifier. As depicted, the accuracy was 100 % both for the training (i.e., validation in training– 222 observations) and testing (55 observations) data.

As the next step, selected validated ML models were used to predict/identify PD sources in the service-aged (50 years) stator winding insulation. The same data

Table 2 – Summary of the accuracies of selected ML classifiers. k: no. of neighbors, γ: Kernel scale.

Classifier	Hyperparameters	Accuracy
Coarse kNN	k=100, metric: euclidean, weight: equal	69.8 %
Cosine kNN	k=10, metric: cosine, weight: equal	97.7 %
Fine kNN	k=1, metric: euclidean, weight: equal	99.5 %
Optimized kNN	k=1, metric: spearman, weight: squared inverse	100.0 %
Coarse Gaussian SVM	multiclass method: one-vs-one, γ= 23	82.0 %
Fine Gaussian SVM	multiclass method: one-vs-one, γ= 1.4	94.1 %
Linear SVM	multiclass method: one-vs-one, γ: auto	99.1 %
Optimized SVM	Function: Cubic, multiclass method: one-vs-one, γ=18	100.0 %
Boosted tree Ensemble	method: AdaBoost, max. splits: 20, no. of learners: 30	69.8 %
Bagged tree Ensemble	method: bag, max. no. of splits: 221, no. of learners: 30	97.7 %
Optimized Ensemble	method: bag, max. no. of splits: 49, no. of learners: 11	100.0 %
Fine tree	max. splits: 100, split criterion: Gini's div. index	96.4 %
Optimized tree	max. no. of splits: 5, split criterion: max. dev. reduction	97.3 %

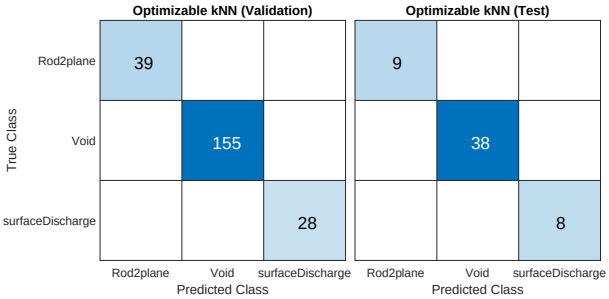


Fig. 8 – Confusion matrix for the optimized kNN-classifier for training (validation) and test.

extraction methodology (de-noising, feature extraction, and data normalization/standardization) was followed before feeding the model with the data. PRPD data from three different service-aged winding insulation was acquired at their rated voltage (7.4 kV). 25 observations from each bar were obtained, amounting to a total of 75 observations where each measurement/observation incorporated 1000 AC cycles at 50 Hz (20 s). As previously mentioned, the stator bars did not contain any visible damage, nor did they have end-windings. Thus, a human expert would expect to observe PDs reminiscent of internal voids and delaminations. And based on the PRPD plots of the winding insulation, as shown in Fig. 5(d), they would label the PD source as internal discharge. However, we had not labeled the observations as "void" but as "unknown." Table 3 shows the predicted PD classes in the winding insulation by the selected models. The tuned SVM model classified 74/75 observations as "void," which was the expected response. Only one observation seems to be misclassified. The

optimized ensemble model yielded similar results, while the rest of the models predicted fewer voids.

As a caveat, ideally, a well-defined PD source in a different experimental setting would have been more assuring to test the model's accuracy, even though the initial results seem to be promising and indicate that the identification of different PD sources within a test object is viable. In future work, further tests will be performed on winding insulations with other types of known defects to endorse the models' accuracies more reliably. Then, the selected models will be tested with the unseen labeled experimental data representing each PD class. Moreover, not all the 32 features used in this work are likely to have the same importance for the trained models because some of them are close descendants of each other, and some may have random values that might mislead the training procedure and cause over-fitting problems. After further endorsing the models using labeled unseen data, feature selection methods (i.a., neighborhood component analysis, MRMR, Chi2, ANOVA), dimensionality reduction techniques (i.a., principle component analysis, multidimensional scaling), and seeking different features will be performed to see if the models can be represented with fewer features without compromise in the overall accuracy. An overview of various qualified models will also be given based on the selected features. All pertinent ML classifiers, including neural networks, will be tested.

Last but not least, new classes to represent other types of PD sources will be included as well as defining different types of voids studied in this work as separate classes, and then the classifiers will be tested under those circumstances.

Table 3 – Predicted PD sources in the stator bars (25 observations from each of three winding insulation).

Model	Rod2plane	Void	surfaceDischarge
Optimized SVM	1	74	0
Optimized Ensemble	3	72	0
Optimized kNN	7	68	0
Optimized Tree	49	24	2
Boosted tree Ensemble	49	13	13

4. Conclusions

The main aim of this work was to test out several ML-based models to classify different PD sources accurately, investigate the benefit of analytics for condition monitoring of HV insulation, and address challenges and knowledge needs for future studies. Lab-made artificial defects were used to generate $\phi - Q_a - n$ datasets from which novel features were extracted. The generation of high-quality training datasets was a success, and they will be extended to incorporate more PD sources, especially those arising in stator winding insulation. Several ML classifiers were trained based on the extracted features, and validation accuracies thereof were benchmarked. Several types of SVM, ensembles and kNN models achieved significantly high accuracies of PD defect identification ($\geq 95\,\%$). Then, the selected classification models were used to predict possible defect types in service-aged winding insulations. The chosen models predicted the dominating PD source in the stator winding insulation to be of void discharge type, agreeing with the initial expectation. However, further tests should be performed on winding insulations with other known defects to more reliably verify the models' accuracy.

This study has laid the groundwork for the future investigation of feature selection and reduction, as well as the introduction of novel statistical features that are fingerprints of PD data. Finding the right features that adequately define and classify different PD sources is still ongoing for further improvement. In particular, five parameter Weibull fit approach yielded promising results and deserves further attention to identify simultaneously active PD mechanisms, e.g., slot discharges and internal discharges in the winding insulation.

Acknowledgment

This work is funded by Digital Twin Laboratory for Hydropower Project through HydroCen FME Centre of Excellence (https://www.ntnu.no/hydrocen/twinlab).

References

[1] CIGRE WG A1.10, "TB 392 Survey of Hydrogenerator Failures," *e-cigre*, 2009.

[2] C. Sumereder, "Statistical lifetime of hydro generators and failure analysis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 15, no. 3, pp. 678–685, Jun. 2008.

[3] M. Istad, M. Runde, and A. Nysveen, "A Review of Results From Thermal Cycling Tests of Hydrogenerator Stator Windings," *IEEE Trans. Energy Convers.*, vol. 26, no. 3, pp. 890–903, Sep. 2011.

[4] M. Levesque, N. Amyot, C. Hudon, M. Bélec, and O. Blancke, "Improvement of a Hydrogenerator Prognostic Model by using Partial Discharge Measurement Analysis," Aug. 2018.

[5] S. Lu, H. Chai, A. Sahoo, and B. T. Phung, "Condition Monitoring Based on Partial Discharge Diagnostics Using Machine Learning Methods: A Comprehensive State-of-the-Art Review," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 27, no. 6, pp. 1861–1888, Dec. 2020.

[6] International Electrotechnical Commission, "IEC 60270:2000+AMD1:2015 CSV — High-voltage test techniques - Partial discharge measurements," Tech. Rep., 2015.

[7] "IEC/IEEE Guide for the Statistical Analysis of Electrical Insulation Breakdown Data (Adoption of IEEE Std 930-2004)," *IEC 62539 First Edition 2007-07 IEEE 930*, pp. 1–53, 2007.

[8] M. Cacciari, A. Contin, and G. Montanari, "Use of a Mixed-Weibull Distribution for the Identification of PD Phenomena [Rotating Machines]," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 4, pp. 614–627, Aug. 1995.

[9] T. Tanaka and T. Okamoto, "Analysis of q-n and φ-q characteristics of partial discharge in several electrode systems," in *1980 IEEE International Conference on Electrical Insulation*. IEEE, 1980, pp. 190–193.

[10] E. Gulski, "Computer-aided measurement of partial discharges in HV equipment," *IEEE Transactions on Electrical Insulation*, vol. 28, no. 6, pp. 969–983, 1993.

[11] E. Eberg, T. G. Aakre, G. Berg, and S. Hvidsten, "Comparison of Offline VLF PD Measurements and Online PD Measurements on a 50-Year-Old Hydrogenerator Stator in Norway," in *2018 IEEE Electrical Insulation Conference (EIC)*, Jun. 2018, pp. 542–546.

Vedlegg 3

SINTEF

Proposal for Knowledge-building Project for Industry (KSP) **2022**

HighFlex

High voltage generators for flexible hydropower

The idea

Hydropower's role and importance is ever increasing as more intermittent renewable energy production is phased into the European power grid. Large-scale wind energy deployment is expected in the Nordics due demands for CO_2 emission reduction simultaneously to increased energy demands. This will result in the need for a much more dynamic and flexible operation of hydropower plants, which will increase the imposed stresses on a large scale. The effect of this on the hydro generator's health and expected lifetime has not been adequately addressed – especially for older hydro generators.

International statistics show that the insulation system is most prone to faults and also is responsible for the longest outage times for hydro generators[1]. A robust, reliable, and user-friendly non-destructive diagnostic test methodology is needed to reliably assess the current state-of-health of electrical systems, and assess if the ageing insulation systems and supporting structures can withstand future dynamic loading, and hence mitigate unexpected failures.

The main goal of this project is to develop testing methodologies, physical/ hybrid/ data driven models for **accurate health prediction of dynamically operated** hydro generators in a grid with increased penetration of intermittent renewable energy sources, especially wind energy – thus **maintaining the reliability of operation**, **security-of-supply** and **improved utilization** of assets.

R&D-activities

- Establish knowledge of the effect of the dynamic operation on relevant hydro generator generations and insulation technologies through theoretical and experimental investigations
- Develop cost-efficient holistic health monitoring methodology for hydro generators' electrical systems, considering both off-line and on-line (periodic /continuous) methods
- Make a structured open database of data retrieved from electrical insulation systems of hydro generators in laboratory and in-service available for the hydropower industry and research community
- Develop models and tools for reliable assessment criteria by applying machine learning algorithms for easy application directly on raw data
- Develop models and provide accurate input for cost/benefit models of dynamic operation in hydropower

[1] CIGRE Technical Brochure 392, 2009.

- Educate MSc- and PhD-candidates with relevant competence for the Norwegian hydropower industry

Value Potential

Improved health prediction and knowledge on degradation under dynamic operation will provide service providers, generator owners, and other utilities with a better decision basis for executing efficient maintenance and asset management:

- Extend the lifetime of generators or plan for optimal refurbishment/replacement
- Provide better cost estimates for facilitating flexibility to the grid
- Reduce the number of physical inspections and repairs, thus reducing the probability of accidents. Generic results also apply to other HV components, such as large wind turbines currently operating at 3.3 kV

Knowledge on degradation under dynamic operation will contribute to other stakeholders including wind power, manufacturing industries, and electrification of transport (ships, road, aviation):

- Evaluate the cost related to increased degradation related to increased flexibility - e.g., for LCA of wind power installations
- Provide insight on degradation/lifetime in high voltage rotating machines

Deliverables

- Physical, hybrid and data-driven methods and models for health prediction
- Knowledge and models on failure and degradation of hydro generators under extreme stresses
- Method for cost-efficient off-line assessment
- Holistic sensor system proposal for on-line assessment
- Knowledge on the cost of dynamic operation and lifetime reduction

Sustainability

HighFlex will contribute to the increased penetration of renewable energy sources and further exploit possibilities for hydropower as an efficient, clean alternative for regulating power into the grid.

HighFlex will contribute to optimal utilization of raw materials, e.g. copper and iron, in existing and new hydropower plants, further reducing the CO_2 footprint.

Budget

Total 20 MNOK (2 M€) over four years.

Financing

With 75% funding from the Research Council and for example five partners, each partner would contribute with 200 000 NOK (€ 20 000) per year.

Partners

R&D: SINTEF Energy Research, Norwegian University of Science and Technology (NTNU).

Potential partners: Renewable energy utilities and industry.

Timeline

7. Jan.: Teams-meeting for potential partners

1. Feb.: Letter of intent

9. Feb.: Application deadline

June: Project start

More information

For more information about the project proposal, please contact:

SINTEF:
Espen.Eberg@sintef.no / +47 915 40 305
Emre.Kantar@sintef.no / +47 934 77 795

NTNU:
Frank.Mauseth@ntnu.no / +47 735 94 234