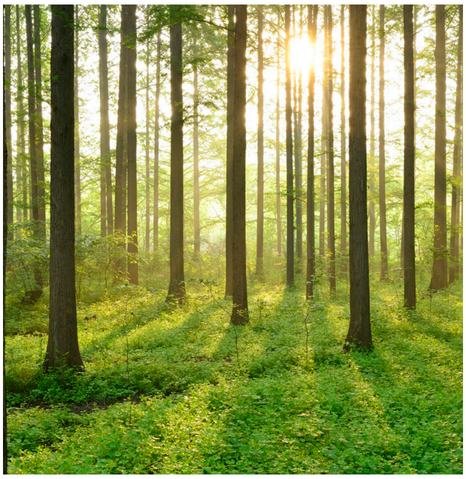
WIND TURBINE PERFORMANCE DECLINE IN SWEDEN

REPORT 2017:436



VINDFORSK

Energiforsk

Wind turbine performance decline in Sweden

JON OLAUSON & MIKAEL BERGKVIST, UPPSALA UNIVERSITY
PER EDSTRÖM, SWECO
NILS-ERIK CARLSTEDT, VINDSTAT

ISBN 978-91-7673-436-0 | January 2018
Energiforsk AB | Phone: 08-677 25 30 | E-mail: kontakt@energiforsk.se | www.energiforsk.se

Foreword

"Wind turbine performance decline" is a project funded by Energiforsk and the Swedish Energy Agency within the program Vindforsk.

How much wind turbine performance declines with age is a question of great importance for the profitability of wind farms. In the long run it affects the required installed capacity to fulfil renewable energy targets. Previous large-scale studies of this type are very scarce why this report is an important contribution for investors and owners of wind turbine facilities.

The project was carried out by Jon Olauson, Uppsala University, Mikael Bergkvist, Uppsala University, Per Edström, Sweco and Nils-Erik Carlstedt, Vindstat

The reference group consisted of Åsa Elmqvist, Energiforsk, Jan-Åke Dahlberg, Vattenfall and Martin Löfstrand, OX2.

Göran Dalén
Chairperson, Vindforsk

Sammanfattning

I denna rapport studeras prestandaförändringar hos vindkraftverk i Sverige. Under de första åren efter idrifttagning är produktionen närmast konstant, men sedan börjar den minska. Verk byggda före 2007 tappar i snitt 0,15 procentenheter per år i absolut kapacitetsfaktor, vilket motsvarar en energiförlust om ca 6 % över livslängden. En gradvis ökning av stilleståndstiden står för 1/3 av minskningen och försämrad effektivitet för den resterande delen. I jämförelse med resultat från Storbritannien tappar svenska verk betydligt mindre prestanda.

I princip alla tekniska system har någon slags försämring av prestandan över tid och det bör därför inte komma som någon överraskning att vindkraftverk producerar sämre mot slutet än i början av sin livslängd. Trots att mycket forskning har gjorts om t.ex. stilleståndstid och havererande komponenter finns förvånansvärt få arbeten om den generella produktionsminskningen över tid. En välgjord studie av de engelska forskarna Staffell och Green utgör det huvudsakliga undantaget. Det huvudsakliga målet med denna rapport är att studera om svenska vindkraftverk tappar lika mycket i prestanda som de i Storbritannien.

Som underlag för studien användes produktionsmätningar från två olika källor: månadsdata från Vindstat och timdata från Cesar (elcertifikatsystemet). Efter att ha rensat bort ca hälften av mätserierna, framförallt mätningar kortare än fem år, återstod 1100 resp. 1300 tidsserier som användes i analyserna. Anledningen till att mätningar kortare än fem år inte beaktades var att inslaget av slumpen i de beräknade trenderna då bedömdes bli för hög.

De studerade verkens konstruktionsår var mellan 1984 och 2010 medan datainsamlingsperioden var 1990–2015. Rådatat har gåtts igenom i detalj och i förekommande fall justerats. Några viktiga exempel på fall där data justerades är:

- På grund av ekonomiska svårigheter har en del verk justerat ner effekten till 1500 kW under senare år. En del verk har också höjt effekten genom mjukvaruuppdatering. Då dessa fenomen önskades exkluderas från beräkningen av trender användes en transformering för att räkna om tidsserierna till en konstant maximal effekt.
- I vissa fall ändrades antalet verk som var kopplade till en Cesar-mätare under mätperioden. Dessa tidsserier delades då upp i två delar.
- En mycket liten del av den totala datamängden var uppenbarligen felaktig och rensades bort.

En korskörning av 97 mätserier från Vindstat, Cesar samt SCADA-data från Vattenfall visade på vissa skillnader i enstaka fall, men på det stora hela bedöms kvaliteten på materialet som hög.

Eftersom vindklimatet varierar på både kort och lång sikt genomfördes normalårskorrigeringar med data från tre meteorologiska modeller: MERRA, ERA-Interim och ConWx (CFSR och JRA-55 undersöktes också men bedömdes vara mindre lämpliga för ändamålet). Till skillnad från den engelska studien

analyserades inte trender i icke normalårskorrigerad produktion överhuvudtaget. Även om det inte är huvudsyftet med denna rapport gör det stora antalet studerade verk att en jämförelse av produktionsmätningar och produktion modellerad från meteorologiska data kan ha ett intresse i sig själv. Vi visar att mediankorrelationen i timvisa data är 0.86 för både MERRA och ConWx och något lägre (0.83) för ERA-Interim. För månadsvärden är motsvarande siffror 0.95 och 0.96 för MERRA och ConWx. Även i andra studerade hänseenden presterar MERRA och ConWx likvärdigt medan ERA-Interim är något sämre.

Tre olika statistiska metoder användes för att beräkna linjära trender i månadsvisa kapacitetsfaktorer (KF) som funktion av ålder. Den analysmetod som mest fokus lades på var linjär regression för enskilda vindkraftverk. Resultaten för de två andra metoderna (linjär regression för ”kohorter” med liknande startår samt den egenutvecklade metoden ”ekvivalent trend”) är dock mycket likartade, vilket visar på en robusthet i analysen. Alla analyserna utfördes på de sex möjliga kombinationerna av mätning och data för normalårskorrigering; Cesar/MERRA, Vindstat/ConWx etc. Resultaten för dessa sex olika kombinationer skiljde sig något mer än de för de olika statistiska metoderna, men inte mer än att tydliga mönster och statistiskt signifikanta resultat kunde erhållas.

Med linjär regression för varje enskilt verk är KF-medianminskningen 0,10 procentenheter per år (pe/år). För verk tagna i drift innan 2007 är median-minskningen 0,15 pe/år, vilket exempelvis motsvarar en försämring av KF från 27 % till 24 % över 20 år. Detta innebär att livstidsproduktionen är 6 % mindre än om verken producerat lika bra som när de var nya. Konfidensintervall för medel- och mediantrender beräknades med en ”bootstrap”-metod eftersom trenderna inte är normalfördelade. Det går med god marginal att påvisa att de negativa trenderna är statistiskt säkerställda.

Eftersom de flesta verk har mindre branta trender de första åren är det mycket som talar för att resultaten för de äldre verken (drifttagna 2007 eller tidigare) är mer representativa sett över hela livslängden. Med andra ord är vår bedömning att även nyare verk kommer att börja tappa prestanda när de åldras även om så inte skett hittills. Huruvida denna bedömning är korrekt kommer vi dock säkert veta först om 10 - 15 år.

I och med den unika databasen med timvisa mätningar var det möjligt att identifiera perioder av stillestånd hos verken (som inte beror på låg vindhastighet). I genomsnitt är stilleståndstiden 4,0 %, men värdet ökar med åldern; från ca 3,2 % när verken är relativt nya till närmare 6 % vid åldern 14-19 år. Denna ökning kan förklara ca 1/3 av den observerade produktionsminskningen. För verk som tappar mycket i prestanda bidrar ofta ökat stillestånd till en stor del av minskningen. Vi visar också att stilleståndstiden generellt är större under vintermånaderna, speciellt för verk i norr.

Genom ett par olika varianter av multipel linjär regression studerades huruvida olika faktorer såsom startår, terrängtyp, fabrikat och ägandeförhållande har en statistiskt signifikant påverkan på trenderna. Valet av analysmetod berodde på att många av faktorerna är relativt starkt korrelerade (t.ex. är nyare verk ofta byggda i skogsmiljö) och en analys av faktorerna var och en för sig skulle därför kunna ge missvisande resultat. För kategoriska variabler som terrängtyp och tillverkare behöver en ”baskategori” väljas och de övriga kategorierna jämförs sedan mot denna. Ingen annan faktor än startår visade på statistiskt signifikanta resultat för alla olika analyser som gjordes (som nämnts tidigare har äldre verk mer negativa trender). Det finns dock relativt tydliga indikationer på att:

- Verk med högre KF har en mer negativ trend.
- Verk i skog har en mindre negativ trend än verk i slättlandskap (baskategorin).
- Verk från ”övriga tillverkare” tappar mer i prestanda än Vestas-verk (baskategorin). Övrigt-kategorin består av verk med annan tillverkare än Vestas, Enercon eller WindWorld. Eftersom de nyaste studerade verken har startår 2010 finns inte så många turbiner av t.ex. GE, Siemens och Nordex med i övrigt-kategorin; den består framförallt av verk som inte längre säljs på den svenska marknaden.

En viktig aspekt med detta arbete är att kunna ge rekommendationer för vindenergi-beräkningar för nuvarande och framtida projekt. Vår uppskattning är att antagen försämring av KF bör ligga i intervallet 0,10–0,20 pe/år. Den mer optimistiska skattningen 0,10 motsvarar medianen för alla studerade verk, 0,15 motsvarar medianen för verk drifttagna 2007 och tidigare (som alltså har producerat under en längre tid) och den konservativa uppskattningen 0,20 härrör från ett påslag för att verk med högre KF tycks ha något skarpare minskning i absoluta tal.

För verk med KF 0,30 och 0,40 motsvarar en nedgång med 0,10–0,20 pe/år ett 20-årigt (år noll till år 19) energitapp om 3,2–6,3 % resp. 2,4–4,8 %. Dessa nivåer är högre än vad som vanligen antas i vindkraftbranschen idag, men betydligt mindre än vad resultat från Storbritannien visade. Det är viktigt att påpeka att de rekommenderade värdena inkluderar både ökad stilleståndstid och försämrad effektivitet. Man bör därför använda en antagen stilleståndstid som svarar mot början av livslängden (runt 3 %) snarare än medelvärdet, annars dubbelräknas inverkan från ökad stilleståndstid.

I den här rapporten visas tydligt att svenska vindkraftverk tappar prestanda över sin livslängd. Spridningen mellan olika verk och parker är dock stor; vissa har en mycket brant försämring medan andra tickar på utan nämndvärda problem. Det första steget i att hitta förklaringar och möjliga förbättringsåtgärder har tagits i och med analysen av mönster i vilka verk som tappar mer än andra. Denna visar dock att skillnaderna i huvudsak är stokastiska givet de förklaringsvariabler som funnits tillgängliga. Vi vill därför starkt rekommendera att branschen tar upp stafettpinnen och fördjupar arbetet: varför är skillnaden i försämringen så stor mellan olika parker och vad kan göras för att hela Sveriges vindkraftsflotta ska prestera lika bra som de bättre parkerna?

Summary

In this report, performance trends of Swedish wind turbines (WTs) are analysed. During the first years of operation, the production is nearly constant, but subsequently it begins to decline. WTs constructed before 2007 lose around 0.15 capacity factor percentage points per year in absolute terms, corresponding to a life-time energy loss of 6%. A gradual increase of downtime accounts for around 1/3 of the decline and worsened efficiency for the rest. In comparison to results from the UK, Swedish wind farms deteriorate much slower.

Two different measurement sets were used in the study: monthly data from Vindstat and hourly data from Cesar. After the removal of around half of the time series, mainly due to measurement lengths shorter than five years, 1100 plus 1300 time series remained to be analysed. WT construction years were in the range 1984–2010 and the data recording period 1990–2015. The raw data were reviewed in detail and, if appropriate, adjusted. Validation of 97 time series from Vindstat, Cesar and SCADA-data from Vattenfall revealed small differences for some WTs, but the data quality can, on the whole, be considered high. Since the wind speed varies on both short and long time scales, long-term correction (LTC) of the measurements were performed using data from different meteorological models.

Three different methods were employed for calculating linear trends in capacity factors (CFs) versus age. Even if the results differ somewhat depending on the combination of dataset, LTC data and statistical method, the overall agreement is good, illustrating the robustness of the analysis. Based on linear regression for all individual WTs, the median CF decline is 0.10 percentage points per year (pp/y). For WTs deployed before 2007, the median trend is -0.15 pp/y, corresponding to a CF reduction from e.g. 27% to 24% over 20 years. Since all WTs have less steep trends during the first years of operation, it is reasonable to assume that the latter figure is a better estimate for the lifetime performance. Confidence intervals for mean and median trends were computed with a bootstrap method since the trends are not normally distributed. It is, with a good margin, possible to demonstrate that the negative trends are statistically significant.

The unique dataset of hourly measurements enabled us to identify periods of downtime due to technical issues. Downtime was estimated at 4.0% for all data, but the share increase significantly with age. This increase contributes to roughly 1/3 of the observed performance decline. For WTs with a sharp deterioration, increased downtime often contributes to a large share of the decline.

By using multiple linear regression techniques, it was studied whether some factors impact the trends in a statistically significant way. It was found that no other factor than start year could fulfil this criteria for all analyses performed. There are however relatively clear indications that WTs with higher CFs have more negative trends, that forest farms decline less than farms in open terrain and that WTs from other manufacturers than Vestas, Enercon or WindWorld lose more performance than Vestas WTs.

An important aspect of this work is to give recommendations for wind energy calculations for current and future projects. Our best estimate is a decline in the range 0.10–0.20 pp/y including both increased downtime and worsened efficiency. We recommend that a value closer to the higher end should be used for high-CF turbines. For WTs with as-new CFs of 0.30 and 0.40, a decline of 0.10–0.20 pp/y corresponds to a 20-year (year zero to 19) energy loss of 3.2–6.3% and 2.4–4.8% respectively. These levels of energy losses are higher than normally assumed in the wind sector today. As compared to results from the UK, the decline is however considerably smaller.

Table of contents

Abbreviations

ACF	Auto-correlation function
AICc	Bias-corrected Akaike's information criteria
CF	Capacity factor
CF_C	Long-term corrected CF with ConWx data as reference
CF_E	Long-term corrected CF with ERA-Interim data as reference
CF_{LTC}	Long-term corrected CF
CF_M	Long-term corrected CF with MERRA data as reference
ConWx	Company providing weather services, e.g. data for long-term correction
ERA	ECMWF (European Centre for Medium-Range Weather Forecasts) Re-Analysis
LTC	Long-term correction, Long-term corrected
MERRA	Modern Era Retrospective-Analysis for Research and Applications
PC	Power curve
pp/y	Percentage points per year
SCADA	Supervisory control and data acquisition
SG14	Study by Staffell and Green [1]
WI	Wind index (a LTC method)
WT	Wind turbine

1 Introduction

Practically all technical systems are subject to deterioration of some degree [2], and it should come as no surprise that the performance of wind turbines (WTs), like e.g. gas turbines [3], decline over time. Although lot of research has been conducted on e.g. downtime and failing components, surprisingly little work has been devoted to the overall reduction of wind turbine performance with age; the thorough study on UK farms by Staffell and Green [1] (SG14) is the main exception. The primary objective with this report is to reproduce this study for Swedish conditions. We however use partly different (improved in some cases) methods and expand the scope by also considering factors explaining differences in trends and the evolution of downtime with age. The latter was made possible by a unique database of hourly observations for almost all Swedish wind turbines.

The main conclusion from SG14 was that UK farms lose around 1.6% of the output per year (in relative terms), with average capacity factors (CFs) declining from 0.285 when new to 0.21 at age 19. This decline is much steeper than what is accounted for in energy calculations performed before the deployment of wind farms in Sweden. If a similar pattern was to be seen for Swedish conditions, the assumptions would have to be changed considerably. Needless to say, this would have a profound impact on the profitability and may ruin the business case for many projects. As will soon be shown, Swedish farms have more gentle performance declines, but our results still call for a change of the assumptions used in the wind community.

This report is structured in the following manner. A literature review is given in Section 2. In Section 3, the databases of wind power time series as well as the processing of the raw data are described. This section also contains a description and analysis of meteorological reanalyses used for long-term correction (LTC) of measurements. The different methods that were used are presented in Section 4 and results are given in Section 5. In order to facilitate reading the report, some material is presented in an Appendix. This introduction section is finished with some clarifications regarding terminology and concepts.

We refer to the entity for which a measurement is taken as a “unit”. In most cases, a unit corresponds to a single WT, but it can also be a small farm. Two different datasets of wind power generation were used in this study and each of these were long-term corrected with three different reanalyses. We thus have six possible combinations of datasets/reanalyses, which we refer to as “cases”. It is often desirable to study trends of units with similar start years. A cohort is, if nothing else is specified, all units for a given case with start years in a three-year sliding window. As an example, the cohort 1995 consists of units with start years between 1994 and 1996. Note that the performance trend for a cohort can be computed either as e.g. the median of slopes for individual units in the cohort or as the slope from a regression of all CFs versus age.

“Trend” and “slope” are used interchangeably to describe the linear change of long-term corrected CF versus age. In other words, the β_1 in $CF_{LTC} = \beta_0 + \beta_1 \cdot age + \ldots$ where age is measured in years (the regression equation sometimes also contains other terms, hence the "..."). The trend/slope is presented in percentage points per year (pp/y) in order to see the results more clearly, e.g. -0.20 pp/y if the fitted model is $CF_{LTC} = 0.31 - 0.0020 \cdot age$. The CF is however generally presented as a dimensionless number (not in percent). Note that the slope above, like generally in this report, represents changes in the absolute CF. Slopes in relative CFs (normalised to as-new CFs) are sometimes also considered.

Two potential terminology pitfalls can be mentioned. When discussing new/old units, we refer to the start year and not the age. In other words, an old unit may e.g. be built in the 1990s and contains both low and high age observations. Also pay attention to the difference between a performance reduction of x% over 20 years and an energy loss of x% over 20 years. The former means that the output at age 20 is x% lower than when new, the latter that the total energy production over 20 years is x% lower than it would be had it been no performance reduction.

Finally, age is measured in years: 0 for the first operational month, 1/12 for the second, etc. When referring to age in integer years, age 3 for example represents the average of monthly observations for ages between 3 and 3+11/12.

2 Literature review

Per definition, a trend in the properly long-term corrected[1] CF of a wind farm can stem from a trend in downtime and/or a trend in the efficiency of the plants. A negative trend could thus be attributed to i) more frequent and longer stops due to technical failures and maintenance and ii) a gradual deterioration of the power curve due to e.g. worsened aerodynamic performance and increased friction in the mechanical components.

We begin by reviewing the literature on the combined effect of these two factors, i.e. studies on the overall deterioration of capacity factors. In Section 2.2, some studies on reduced efficiency are presented. Sections 2.3–2.4 deal with the downtime due to technical failures and failing components of WTs respectively. The review is concluded with an examination of some studies on datasets for long-term correction.

2.1 OVERALL DETERIORATION OF CAPACITY FACTORS

The main inspiration for this work is SG14 [1], which were in turn inspired by, and used largely the same dataset as, Hughes [4]. These studies looked at the general trend of wind farm performance in the UK. As pointed out in SG14, there seem to be no earlier studies of long-term loss of output in the open literature. As of November 2016, none of the 70 publications[2] currently citing SG14 presents any new analysis of CF deterioration. Our review of studies on this topic is therefore limited to the above-mentioned works and an informal study for the Danish case by Bach [5]. Given the attention the works by Hughes and SG14 have attracted, it is likely that similar studies will be performed for other regions in the near future.

In SG14, up to eleven years of monthly capacity factors for 282 farms in the UK were analysed. Results were presented for both raw and LTC data, using a few different statistical methods. The linear trend was in average -0.43 pp/year in absolute terms or -1.6 pp/y relative the average UK capacity factor. From Figure 9b in SG14, it seems that the average CF increases sharply during the first 1–2 years of operation, declines relatively slowly during age 2–9 years, declines more steeply during age 9–18 and even more steeply for age 18–20. This pattern is compatible with the results in Figure 10b in SG14, where farms built 2005 and later have lower linear decline rates than older farms.

Two non-peer-reviewed studies of CF decay, neither of them using any standard LTC technique, are presented in Refs. [4] –[5]. In the former, CF decay was calculated for the UK and Denmark using an error components model with fixed effects. For UK onshore farms, the capacity factors were found to decline from a peak of 24% at age 1 to 11% at age 15. By weighting the results by capacity, the CF at age 15 was as low as around 3% (which raises the question whether the method

[1] If a long-term correction is not performed, a trend in the CF of a farm could stem from a long-term trend in the wind climate. Furthermore, if the reference time series used in the LTC is flawed, a spurious trend in the long-term corrected CF time series might be observed.

[2] Search on Google Scholar, 2016-11-21

used is indeed appropriate). In comparison to the fixed effect model in SG14, the deterioration is much more severe. For onshore farms in Denmark, the capacity factor falls from 22% to 18% at age 15. For offshore farms, the CF falls from 39% at age 0 to 15% at age 10. Using a different approach, Ref. [5] got much smaller deterioration rates for the Danish case; for onshore farms the deterioration of normalised CFs was 0.3–0.4 pp/y. Furthermore, this deterioration could be fully explained by the reduced mean wind speed seen during the studied period. As for onshore farms, the steep decline for offshore turbines looks alarming but cannot be found in clusters with the same year of installation. Bach therefore concludes that the seemingly dramatic decline of offshore performance is due to a flaw of the approach of analysing all farms together.

2.2 REDUCED EFFICIENCY

A reduced overall efficiency is one of the two fundamental mechanisms that can explain a possible negative trend in wind farm output. Shin and Ko [6] studied empirical power curves (PCs) derived with the method specified in the IEC 61400-12-2 standard, i.e. using nacelle anemometry. Four years of usable data for eleven WTs in South Korea were analysed. By feeding the empirical PCs with wind speeds from a (constant) Rayleigh distribution, the long-term corrected CFs were obtained for each year. The linear CF trends ranged from -0.98 to +0.34 pp/y in absolute terms. Eight out of the eleven WTs had negative trends.

Dalili *et al.* [7] considered three mechanisms deteriorating the aerodynamic performance of WTs: icing, insect contamination and erosion of the blade surface. Icing of the blades can cause severe production losses and decrease component lifetimes due to imbalanced operation. Insects, or other contamination, can increase the surface roughness near the leading edge of the aerofoil, leading to flow separation and a deteriorated performance. The most common solution to reduce the effects of insects and air pollutants on the blades is to wait for rainfall to wash the blades. Erosion of the blade leading edge by sand and water droplets can also increase the surface roughness and reduce power output.

Soltani *et al.* [8] studied the effects of surface contamination on WT performance by comparing characteristics of a clean blade section to ones with different types of roughness tapes (in a wind tunnel). The main sources of increased surface roughness are stated to be insects, ageing, sand impact and rain contamination. Ageing can give rise to cracks which deteriorates the aerodynamic performance. Results from similar tests were presented in Ref. [9], where it was concluded that leading edge roughness can give rise to considerable reduction of the lift coefficient and change the angle where stall occurs. Contaminants can therefore cause severe degradation of performance, particularly for stall-regulated turbines.

Khalfallah and Koliub [10] performed measurements on operating turbines (100–300 kW) near Hurghada, Egypt. During the nine month test period, significant amounts of dust built up on the blades (no raining occurred during the period, neither were the blades washed), causing a gradual deterioration of the measured power curve through increased drag and reduced lift. At the end of the test period, the “mean power loss” was over 50% for the two stall regulated turbines. The pitch regulated turbine was affected to a much smaller degree (less than 20% loss).

2.3 DOWNTIME

Operation and maintenance can constitute a significant part of the lifetime cost of a wind farm, especially for offshore farms [11]. A review of the O&M costs is not within the scope of this study, since we are primarily interested in the downtime and efficiency of the turbines.

According to Ref. [12], downtime of WTs can be classified as follows:

- High-wind stops
- Temperature related stops (WTs cannot operate at ambient temperatures below minus 15–30°C)
- External grid related stops
- Disturbances (these problems can be solved without on-site labour)
- Technical failures (these are the ones included in the component breakdown in Section 2.4)
- Service
- Icing
- Other, e.g. research.

When comparing downtime data from different sources, it is important to know if all the abovementioned factors are included or if some of them are not. In Finland, technical failures and disturbances accounted for 61% and 29% respectively of the total downtime of 3.0% [12]. For comparison, the same figures for Sweden were between 41–64% and 18–35% for the years 2000 through 2006 [13].

The total downtime due to technical failures is the product of the failure rate and average downtime per failure. In Finland, the downtime increased for turbines older than 14 years although, interestingly, the number of errors did not increase significantly. This might be due to difficulties to find components for older turbines, and that it might be questionable whether it is profitable to perform expensive repair work for old turbines. The sample size of older turbines was however small, so one should draw conclusions with caution.

A commonly used explanation model for the failure rate is the “bathtub curve” [14]–[17]. This means that during the early period of a farm’s life, the failure rate is high (teething issues). This phase is followed by a longer period of random failures at a low rate. Finally, towards the end of the lifetime, failures become more common. The bathtub model seems to be somewhat compatible with the results in SG14. According to Hahn *et al.* [18] it can be expected that the failure rate due to “wear-out failures” does not increase before the 15th year of operation and they state that the technical availability is now as high as 98%. Faulstich *et al.* [17] on the other hand noticed an increase in failure rate after year 11 and gave availability figures in the range 95–99% for European onshore WTs. A way to increase the availability is condition monitoring [19], [20]. This method, which has become more widely used in recent years, allow the operator to reduce overall costs by performing maintenance or replacing parts before failure.

Table 1. Subassembly breakdown of downtime due to technical failures. Some results are digitized from figures in the respective publications.

	Ribrant & Bertling [21]	Stenberg [12]	WMEP [17]	WindStats (from [14])	LWK [8]
Country	Sweden	Finland	Germany	Germany	Schleswig Holstein, Germany
Period	2000–2004	1996–2008	1989–2006	1999–2008	1993–2004
Turbines	527–723	Ca 10–70	Up to 1500	Up to 20,000[3]	158–643
Gearbox	19%	18%	10%	25%	23%
Blades/pitch/hub	9%	11%	16%	17%	13%
Electric system	14%	10%[4]	15%	16%	17%
Generator	9%	9%	10%	16%	11%
Control system	18%	4%	11%	5%	4%[5]
Hydraulics	4%	15%	5%	6%	5%
Yaw	13%	4%	8%	5%	4%
Sensors	5%	7%	6%	3%	-
Brakes	1%	10%	6%	2%	10%
Drive train	2%	-	5%	4%[6]	9%[7]
Other	3%	11%[8]	8%	2%	4%

In conclusion, the average downtime reported in literature is in the range 1–6%, with 3% typically taken as industry standard [22], [23]. There is some evidence that the failure rate increases for turbines older than around 10–15 years, but the increase is not very sharp. The downtime per failure can also be longer for older turbines due to e.g. lack of spare parts. It does however not seem plausible to explain the relatively steeply decreasing trend in capacity factors seen in SG14 solely with a gradual increase of downtime.

2.4 FAILING COMPONENTS

In this section, subassembly breakdowns of the downtime due to technical failures are reviewed. Results from five studies in Sweden, Finland and Germany are compiled in Table 1. In general, the databases which the studies draw upon have flaws of different kinds (e.g. changes in the reporting system) so the results must be interpreted with care. The definitions of the subassemblies are not always clear-cut and can be overlapping. Footnotes are given in case a different terminology is

[3] Only a relatively small (and decreasing) share of the turbines report detailed down-time statistics. The data given is for those with details.
[4] Electric system + grid connection
[5] Converter
[6] Main shaft/bearing
[7] Main shaft
[8] Whereof structure 4% and heating 3%

used in the studies. Some results are digitized by us from figures which gives small additional uncertainties.

The results differ somewhat between the studies, but generally the subassemblies contributing mostly to the downtime are the gearbox, the electric system, rotor/pitch-systems and the generator. Hydraulics constitutes a surprisingly large share of the downtime in Finland, but this might be due to a limited sample size. In Ref. [24], references for further reading on component failures are given.

2.5 DATASETS FOR LONG-TERM CORRECTION

The wind speed time series used for LTC of a measurement can come from meteorological models (reanalyses) or long-term measurements. In recent years, the former has become the standard in the wind energy sector. The time series should preferably have a high correlation to the measurements and be consistent in time, i.e. not contain erroneous trends or step changes. A compilation of some studies on reanalysis performance can be found in Ref. [25] and an overview of available reanalysis datasets is given in Ref. [26].

A conclusion from the reviewed literature [26]–[30] is that no model is consistently best in terms of correlation and consistency. However, it seems that the newer datasets generally perform better than NCEP/NCAR [27], [28] and that downscaling using the WRF model [31] can sometimes improve the results [27].

Liléo and Petrik [28] concludes that the strong upward trends for some NCEP/NCAR and CFSR grid points is probably a signature of inconsistencies in the models. Brower *et al.* [30] found that MERRA had discontinuities and statistically significant trends for several of the considered grid points. A lesson one can learn from the trend analyses in Refs. [26], [28] and [30] is that it is probably wise to use two or more independent datasets for LTC; if the results do not differ too much, they can be considered robust.

3 Data

In this chapter, the different datasets of wind power generation are presented and our processing of the data is described (Section 3.1 and 3.2 respectively). The chapter is concluded with a description and analysis of five reanalyses of potential interest for long-term correction of the raw data. Figure 1 gives a summary of all data-processing steps (including those presented in Section 4).

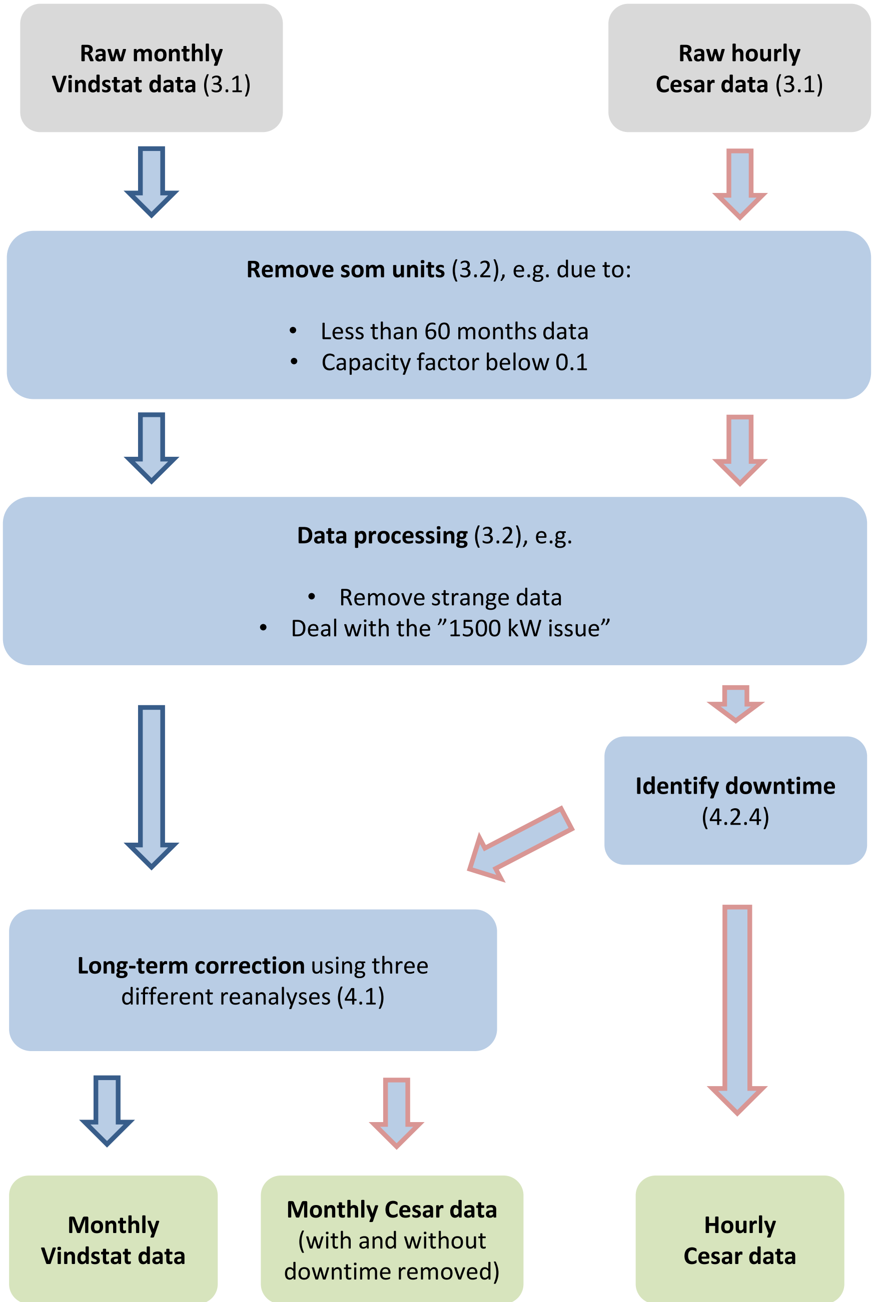


Figure 1. Overview of all data-processing steps. References to sections in this report are given within parenthesis.

3.1 WIND POWER DATASETS

Two datasets were primarily used in this study: Vindstat monthly data[9] and hourly data from the Swedish electricity certificate system Cesar[10]. Due to non-disclosure agreements, the data can unfortunately not be shared.

Energimyndigheten (the Swedish Energy Agency), Elforsk and Vindforsk have financed the collection of data from operational WTs since 1988. Since the practical work is now carried out by the company Vindstat AB, we refer to this database as "Vindstat". Before 2002, the collection was manual, but since 2002 an automatic system is used to collect data for each WT daily. Between 1991 and 2003, an investment support was available for constructing wind farms. A prerequisite for obtaining support was to report generation data to Vindstat. Consequently, almost all older WTs are connected to this system. In 2003, the investment support was replaced by the current electricity certificate system and reporting to Vindstat was no longer mandatory. New WTs are therefore not always included in the Vindstat database; as of the end of 2015, 1945 out of around 3200 WTs were included in the database (3114 out of around 6000 MW). Monthly data from Vindstat between January 1990 and December 2015 were used in the analyses.

The Cesar database contains hourly meter readings for all WTs connected to the electricity certificate system. Since the certificates give an additional revenue of around 15–20 €/MWh[11], almost all WTs are included. Certificates are issued for 15 years except for WTs deployed before 2003, which only got certificates from 2003 to 2012 (2014 in some cases). This implies that the longest time series available in Cesar is around 13 years and that data for older WTs are not available for their early ages. Most often (94% of the time series), each WT is reported separately to Cesar but sometimes production for a whole farm is reported in aggregation. In the continuation of this report, we refer to the entity for which measurements are taken as a "unit". A Vindstat unit is thus always a single WT, but for Cesar it can be either a WT or a small farm. Both hourly and monthly Cesar measurements, ranging from May 2003 to December 2015, were used in the analyses.

In addition to the large datasets from Vindstat and Cesar, detailed data on three farms owned by Vattenfall were available. These time series, containing two years of 10-min data on generation and downtime, were used for validation. Downtime data for nine WTs from OX2 were also available.

Metadata such as coordinates, installed capacities and rotor diameters were compiled from Vindstat reports[12], the Cesar database described above and the electricity certificate system[13].

---

[9] Available from http://vindstat.com/ (Accessed: 2016-11-21). We however used an Excel spreadsheet provided by Nils-Erik Carlstedt, one of the project members.
[10] This dataset was provided by Energimyndigheten (the Swedish Energy Agency).
[11] Currently (Feb 2017), the certificate price is however below 10 €/MWh.
[12] Monthly and yearly reports are available at http://vindstat.com/ (Accessed: 2016-11-21).
[13] Excel spreadsheet of "Godkända anläggningar" obtained from http://www.energimyndigheten.se/fornybart/elcertifikatsystemet/marknadsstatistik/ (Accessed: 2016-11-21).

Regarding the locations of the units, additional sources were sometimes used, including internet searches on particular farms. The accuracies of the locations differ; for many WTs, exact coordinates were available but in some cases, only the centre coordinate of the farm. For other units still, relatively crude coordinates from the Vindstat reports or the location of the nearest village were used. For the purpose of long-term correcting the monthly generation data, all coordinates can however be considered sufficiently accurate.

Rotor diameters and hub heights were not available in the Cesar metadata. A first idea was to try to couple the units in the Cesar and Vindstat databases. Since many Cesar units are not present in the Vindstat dataset, these variables were instead estimated with a machine learning model trained on Vindstat data, see Section 4.3.

3.2 DATA PROCESSING

In this section, the processing of the raw data is described. First, steps common for both datasets are given. Subsequently, unique measures for Vindstat and Cesar are described. Some graphical examples of time series that have been altered are given in Appendix Section 2.1. A comparison of metadata and monthly CFs from Cesar, Vindstat and Vattenfall is given in Appendix Section 3.

The removal of units is a trade-off between quality and quantity of the resulting datasets. Almost half of the units were removed due to measurement periods shorter than 60 months (five years). Lowering this threshold would give more data, but the trends for units with short measurements would be less reliable and meaningful. A more thorough motivation of the chosen threshold is given in Appendix Section 1.1. Apart from units with short measurement periods, a few WTs with installed capacity below 99 kW or average CFs below 0.1 were removed since these were not considered representative for commercial WTs.

When analysing trends in performance, data corresponding to age 0–3 months and higher than 20 years were removed. The reason for not removing all data for the first year (as in SG14) was that WTs generally have full performance after four months. A likely explanation to the difference is that SG14 analysed measurements for farms while we have data for individual WTs in most cases.

For various reasons, the installed capacity can change during the measurement. This is most easily detected with hourly data, but can sometimes also be seen in the monthly time series. Depending on the reason for the change, different actions were taken:

- In the Cesar database, the number of WTs connected to a meter changed in a few instances. These time series were split and each segment was analysed separately (segments shorter than 60 months were however removed).
- Some WTs have a smaller increase in the rating, e.g. from 1800 kW to 2000 kW. We interpret this as a software update which allows a higher output. More common is that turbines have been rated down to 1500 kW around year 2012–2015. The reason for this is the low electricity prices in recent years and that lower fees have to be paid for WTs with rating 1500 kW or lower. For WTs that

have been rated up or down, a transformation was used for the period with higher rating, see Appendix Section 1.2.
- Especially for older units, it is relatively common with long periods (several months) with maximum output at a lower level than the rated capacity. If this lower level is not 1500 kW, we interpret this as down-rating due to technical problems. The time series were thus left without any change.

For both Vindstat and Cesar, problems in the reporting system can cause several months of data to be reported later. Relatively often, the production for the faulty period is given as an average for the whole period. This might not appear as a major problem, but as the example in Figure 2 illustrates, the linear trend of long-term corrected data may change considerably. The solution to this issue is to identify periods with averaged measurements and then to perform the LTC for the whole period in aggregation (i.e. not month by month as is otherwise done).

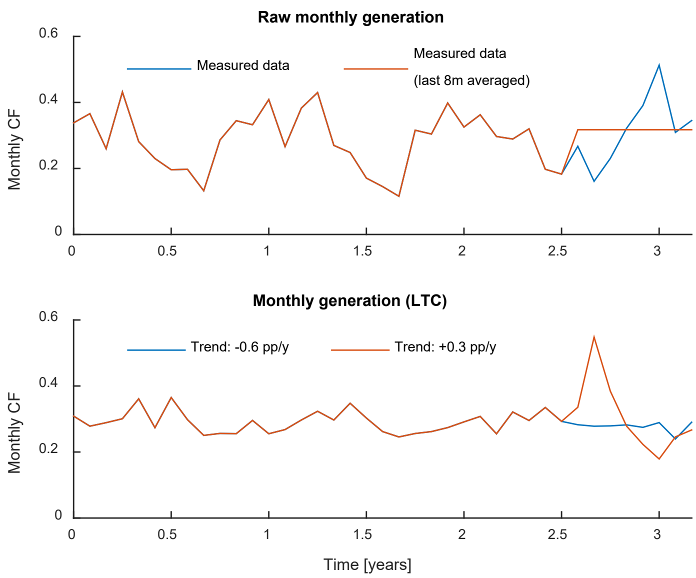


Figure 2. Effect of averaged data on trend estimate. The upper panel shows measured, monthly data during three years and the same time series with the last eight months averaged. In the lower panel, the resulting long-term corrected series are given (LTC of each month separately). The "true" linear trend is -0.6 pp/year, but with averaged data a trend of +0.3 pp/year results. By performing the LTC for these eight months in aggregation, the issue is resolved.

3.2.1 Vindstat

Five prototype units, only operating during a few years in the 90's, were removed from the analysis. A few obviously erroneous capacity factors (below zero or above one) were also removed. For a few units, the reported start date was slightly later than the first month with data. In these cases, the start date was changed.

3.2.2 Cesar

If problems occur in the reporting system so that no measurements are supplied to Cesar, this can be corrected afterwards by adding a "fictive" plant. The time series of the fictive plants are different in nature and were handled differently in the data treatment. Sometimes, the fictive time series are appropriate, hourly data. These were simply added to the original data. In other cases, the fictive time series is a duplicate of already available data and was discarded. Finally, the fictive data can be reported as a mean of a longer period or as a very high value followed by zeros. The total energy is then correct, but the time series cannot be used for analyses of hourly data.

For the around 6% of the units where measurements correspond to several WTs in aggregation, ranges of start dates are sometimes given since it takes some time to erect all turbines. Data for months before the completion of the whole farm were then discarded.

After consultation with the reference group, it was suggested that periods with missing data longer than around one week should be interpreted as turbine downtime. By cross-validating units with long data gaps with the corresponding Vindstat units[14], it was however found that all these had actually generated electricity during the periods of missing data. We thus do not add zeros to these periods.

For some units, zeros are reported at the end of the measurement period. These data can of course correspond to actual downtime, but can according to Energimyndigheten also be e.g. reports from an intermediary actor that does not correspond to downtime. For units with start year before 2003, certificates are no longer obtained after 2012 (or 2014 in some cases). When trailing zeros were present for such units after 2012 or 2014, we removed this data. For newer units, trailing zeros were not removed.

All time series were manually inspected. If the reported installed capacity differed substantially from those seen in measurements (excluding possible peaks due to fictive plants), the installed capacity was changed.

3.2.3 Summary statistics

Table 2 shows the number of units in the two datasets and how many of these that were removed for various reasons. Table 3 shows some statistics for the observations that were used in the final analysis, i.e. after some units and some data were removed. Figure 3 gives the number of units with different start years and the number of observations for different years and ages. Figure 4, finally, shows histograms of observation lengths and monthly capacity factors.

For Vindstat, in average 1.5% of the monthly data is missing (gaps in the time series). For Cesar, the corresponding value is 0.15%. Missing data induces an additional uncertainty in estimation of long-term means [32] and trends, but this is not expected to be a major issue in our case.

[14] 29 Cesar units with gaps longer than one month were studied. Out of these, we were able to identify 22 in the Vindstat database, but only 13 had concurrent data since many units quit reporting to Vindstat in 2003 when the certificate system started. For all 13 units, monthly Vindstat production data was available during the periods with missing data in Cesar.

Table 2. Number of units in the different datasets before and after the removal of units for various reasons.

	Vindstat	Cesar
All units	1990	2755
Experimental units	-5	
Outside Sweden	-2	
<60 months data	-872	-1358
P<99kW		-43
Mean CF<0.1	-7	-41
Split time series		+4
Remaining	**1104**	**1317**

Table 3. Statistics of the datasets used in the analyses, i.e. after removal of short time series, erroneous data etc. Note that for Cesar, the number of wind turbines (WTs) is larger than the number of units.

	Vindstat	Cesar
Number of units	1104	1317
Number of WTs	1104	1537
Total capacity	1.2 GW	1.9 GW
Temporal resolution	Monthly	Monthly / Hourly
Number of observations	143,000	142,000 / 103 millions
Construction years	1984–2010	1988–2010
Data recording period	1990–2015	2003–2015

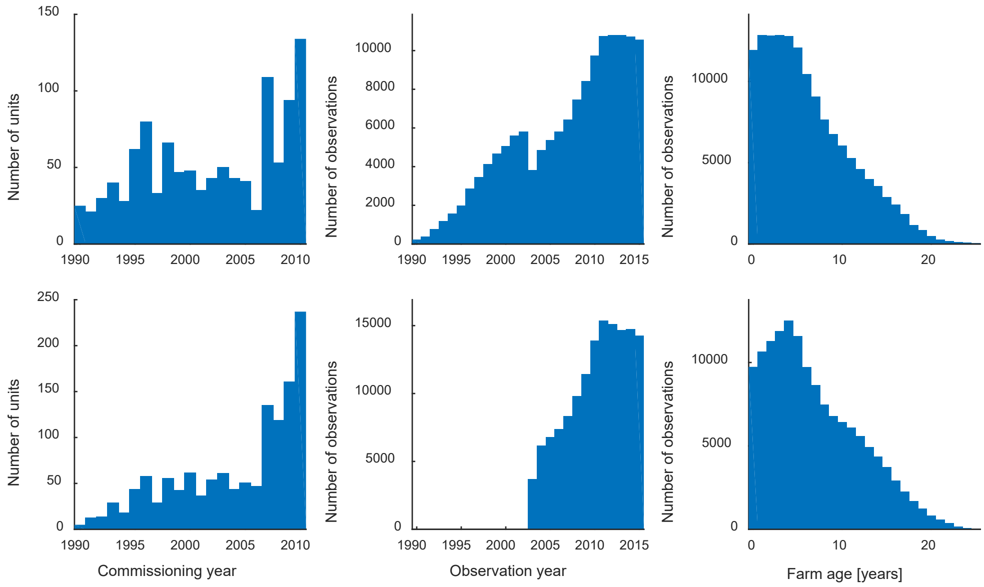


Figure 3. Statistics on monthly data for Vindstat (upper row) and Cesar (lower row). Note the negative step change in Vindstat observation numbers in 2003, which is due to removal of the reporting requirement for obtaining support. The bars for start year 1990 in the leftmost panels also contain a few units with earlier start years.

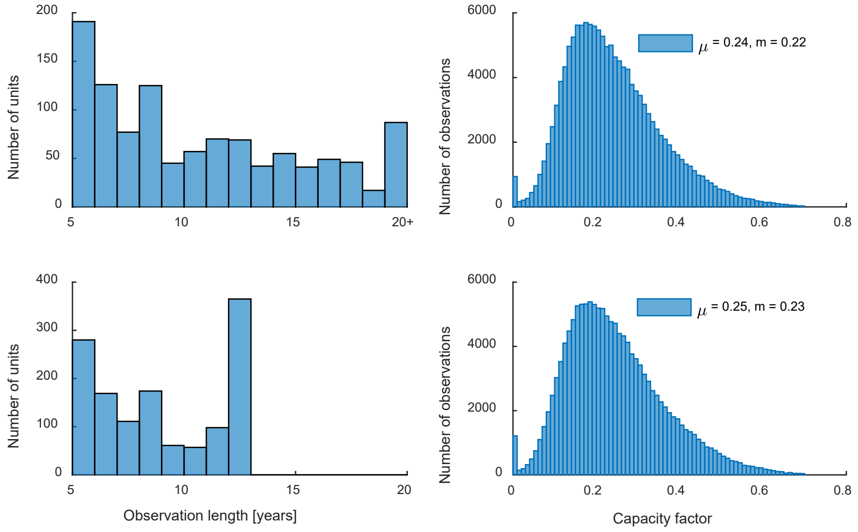


Figure 4. Histograms of observation lengths (left) and monthly capacity factors (right) for Vindstat (upper panels) and Cesar (lower panels). The observation lengths are given as the number of months with data divided by twelve. The mean (μ) and median (*m*) capacity factors are given in the right panels.

3.3 REANALYSES

Four third-generation reanalyses were considered for the long-term correction of monthly measurements: MERRA [33], ERA-Interim [34], CFSR [35] and JRA-55 [36]. The reason for not choosing the new version of MERRA (i.e., MERRA-2) is that Ref. [37] and preliminary studies by us show no improvement over the first generation for Sweden. CFSR data was ruled out at an early stage since the time series are not entirely consistent in time; from 2011 and onwards, only outputs from the operational model CFSv2 are available.

In addition, EMD-ConWx[15] (from now on only ConWx) data for grid points relatively close to all units were downloaded. A maximum distance of 20 km was allowed but the mean distance was 6 km. ConWx, which in contrast to the other datasets is not freely available, is a mesoscale dataset produced by downscaling ERA-Interim. In the following, we for simplicity refer to ConWx as a reanalysis.

Time series from MERRA, ERA and JRA for the whole of Sweden for 1990–2015 were downloaded, i.e. the 26-year period with available wind power measurements. ConWx data was however only available from 1993 and onwards. Furthermore, ConWx is only available for latitudes below 66.5° so a few units were not covered. Some metadata for the reanalyses are given in Table 4.

[15] http://emd.dk/files/windpro2.9/EMDConWx_MesoScale_data.pdf Accessed: 2016-12-22

Table 4. Metadata for the four reanalyses considered for long-term correction of measurements.

	MERRA	ERA-Interim	JRA-55	ConWx
Temporal resolution	1 h	6 h	6 h	1h
Spatial resolution	0.5° × 0.67°	0.75° × 0.75°	0.56° × 0.56°	0.03° × 0.03°
Time period	1990–2015	1990–2015	1990–2015	1993–2015
Heights	50 m	10 m	10 m	25, 50, 75, 100, 150 m[16]

In the following paragraphs, correlations between the reanalyses in terms of monthly output and trends are presented. In Section 5.1, results are given on the performance of the different reanalyses for long-term correction. For each onshore and near offshore point in a regular grid (not coinciding with the native grid of the reanalyses), time series of “fictive” monthly generation for 1993–2015 were calculated. The wind speeds were first bilinearly interpolated in horizontal and linearly scaled to a mean wind speed of 7 m/s. Time series of generation were subsequently calculated using the power curve from a 2.5 MW wind turbine with 100 m rotor diameter. Similar calculations were performed for the 144 available ConWx coordinates.

In Figure 5, the correlations between monthly outputs from the different reanalyses are shown. MERRA, ERA and ConWx are generally in good agreement, but JRA has much lower correlations to the other reanalyses. Average correlations are given in Table 5. Note that the high average correlations for ConWx are partly due to the higher weight on southern grid points (compare upper and lower row in Figure 5).

[16] Data is also available for other heights.

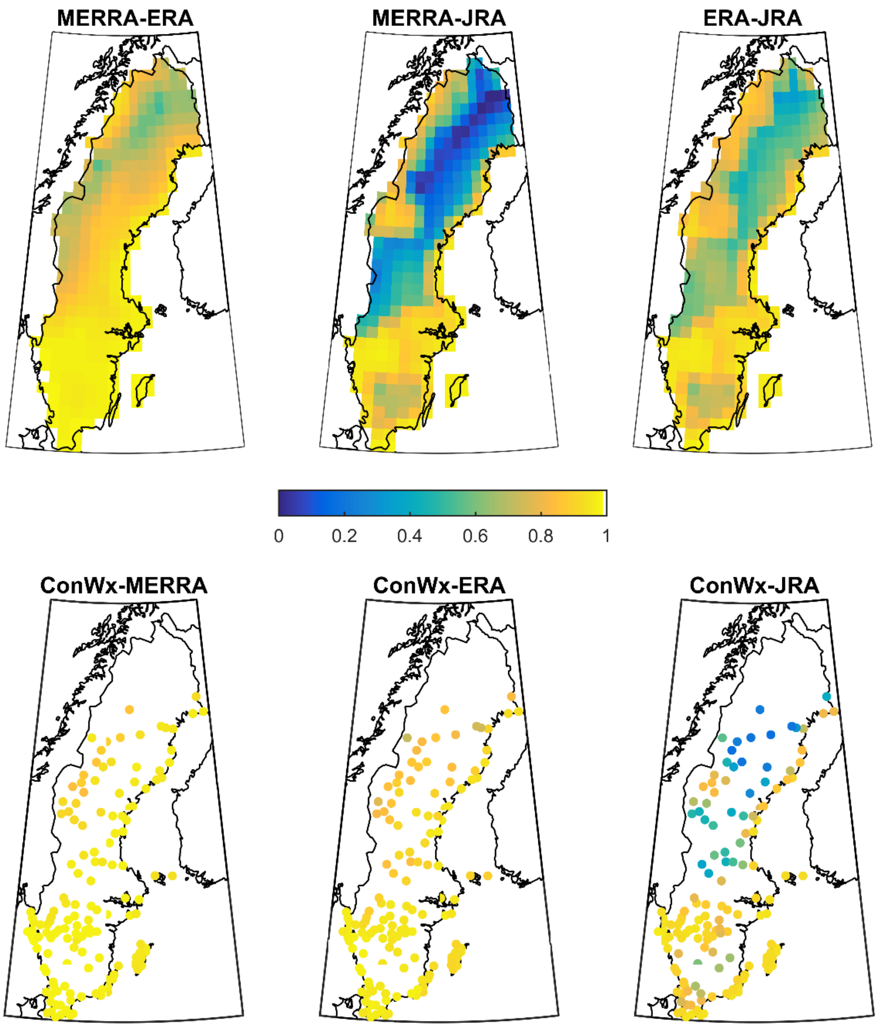


Figure 5. Correlations between fictive, monthly generation time series calculated from the different reanalyses (1993–2015). Note from the lower panels that there are two units on Åland (an island belonging to Finland) in the Vindstat dataset.

Table 5. Average correlations between fictive, monthly generation calculated from the different reanalyses (1993–2015, see Figure 5 for the considered grid points).

	ERA-Interim	JRA-55	ConWx
MERRA	0.85	0.59	0.96
ERA-Interim		0.74	0.93
JRA-55			0.78

Linear trends of monthly generation were subsequently computed for all grid points. The trends for JRA are considerably larger in magnitude and more or less uncorrelated to the trends from the other reanalyses. Trends for MERRA and ConWx are relatively strongly correlated (0.73), but in general somewhat more positive for ConWx. Interestingly, the trends for ConWx and ERA are only weakly correlated (0.32) although ConWx is based on ERA data. A plausible explanation is that ERA is given for 10 m, i.e. considerably lower than the 100 m ConWx level which was used in the calculations above.

From the analyses above and from a preliminary assessment of the correlations between reanalyses outputs and actual measurements, it was decided not to use JRA data. Results based on LTC with MERRA, ERA-Interim and ConWx will thus be given in the remainder of this report.

4 Methods

In this chapter, most of our methods are described. For better readability, some methods are however presented in Appendix Section 1. First, long-term correction is introduced. Second, the methods used for quantifying trends are given. Sections 4.3–4.4 are devoted to machine learning and filtering. The chapter is concluded with a list of differences between our methods and those used in SG14.

4.1 LONG-TERM CORRECTION

In order to remove the effects of variability in the wind climate, long-term corrections of measurements can be performed. If this is not done, artificial trends in wind power performance might be present. A simple but very useful approach, especially for LTC of monthly generation data, is the "Wind Index" (WI) method[17] [39]. For a given unit, the mean wind speed is first estimated, see Section 4.1.2. The reference wind speed time series (reanalysis data in our case) is linearly scaled to this mean. Subsequently, a "fictive" generation time series is computed using the power curve of the WT in question. The WI of a particular month is defined as the average fictive generation of that month divided by the average of the whole time period. By dividing measured monthly generation by the corresponding WI, a long-term corrected time series results, see e.g. Figure 2 on page 22. In the following, let CF_{raw} denote the raw, monthly CF time series. Let furthermore CF_{LTC}, CF_M, CF_E and CF_C denote the long-term corrected CF_{raw} (LTC in general and with MERRA, ERA-Interim and ConWx respectively).

For MERRA and ERA-Interim, the raw wind speeds were taken directly from one model height (see Table 4) and subsequently linearly interpolated to the desired mean. For ConWx, the time series were first interpolated to the hub heights each time step using model outputs from different heights and the power law [40]. In principle, it should be possible to achieve better results for e.g. MERRA by first calculating the hourly wind speed at hub height from 10 and 50 m data before linearly interpolating to the desired mean wind speed; during some hours/seasons, the wind shear is higher than others which could then be captured. If one studies the wind shear calculated from 10 and 50 m MERRA data, this variable is however more or less constant in time, so the results would change very little had the power law first been employed. As a result of neglecting the varying wind shear, seasonal bias sometimes exist for CF_M and CF_E. With ConWx data, on the other hand, the calculated wind shear has a much more realistic pattern with large differences between different meteorological conditions. Despite this, CF_C also has seasonal variations for some units.

In Figure 6, the national, yearly WI between 1993 and 2015 is shown. The WI is calculated as the capacity weighted average of WIs of units in the Vindstat dataset. In SG14, results were given both with and without LTC of the data. Based on Figure 6, analysing the trends on CF_{raw} would not be meaningful for Sweden; the wind climate has changed so much during the last years that strong artificial

[17] Several different WI definitions exist, see e.g. [38].

trends in performance would be present. One can also note that although the different WIs generally agree well, the MERRA data gives somewhat lower values for the last years.

Figure 6. National, yearly wind index calculated as the capacity weighted average of WIs in the Vindstat dataset for 1993–2015.

4.1.1 Power curves

Power curves (PCs) are necessary inputs to the WI calculation. Instead of using actual PCs of all different WTs, a simplified (in terms of workload) approach was taken. Four different PCs for WTs with specific ratings (installed capacity over rotor area) ranging from 229 to 497 W/m² were obtained and normalised to the installed capacity. For each WT, the PC was subsequently estimated by linear interpolation depending on the specific rating. An example is given in Figure 7.

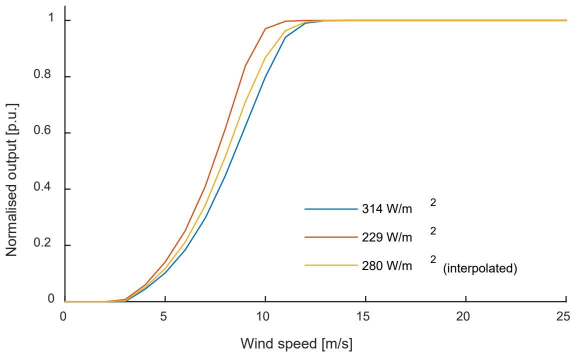


Figure 7. Example illustrating how a power curve for a wind turbine with specific rating of 280 W/m2 is interpolated from two actual power curves.

We thus assume that WTs with similar specific rating have sufficiently similar normalised power curves. A closer examination revealed that the exact shape of the PC has very little impact on the monthly WI as long as the wind speed time series are scaled to give a certain long-term CF, see Figure 8. The assumed CF, however, has a relatively strong effect on the WI variability. A lower CF gives larger variations in the WI and conversely. If, for example, the true long-term CF is underestimated and a trend in the wind climate exists, the LTC will overcompensate for this trend, resulting in an opposite but artificial trend in WT performance. This is why the estimation of the true CF, see next subsection, is so important for trend analyses. We therefore believe that the methodology in SG14 (estimation of true mean wind speeds directly from coarse MERRA data) is inappropriate[18].

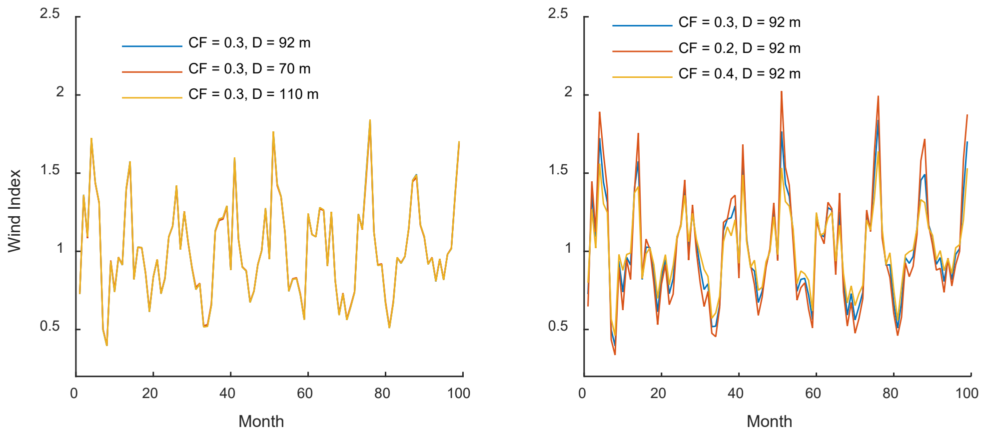


Figure 8. Impact from rotor diameter and assumed capacity factor (CF) on the wind index (WI). As long as the CF is the same, the diameter has negligible impact (left panel). The assumed CF, however, has a profound effect on the WI (right panel); a lower CF gives considerably larger variability.

4.1.2 Estimation of mean wind speeds

As shown in Section 4.1.1, an appropriate estimation of the “true” long-term mean wind speeds is crucial for calculation of the WI. Our take on this matter was to first estimate the long-term CF from the recorded data and subsequently find the corresponding mean wind speed given the WT characteristics and distribution of wind speeds at the actual site.

In order to obtain more robust results, the long-term CF was estimated in two steps. First, the whole time series of monthly measurements was long-term corrected using the mean of CF_{raw} as the long-term CF estimate. Subsequently, the mean of preliminary long-term corrected observations for age 0.33–3.33 was calculated. This is our final estimate of the long-term CF which is used for calculating the WI and corresponding CF_{LTC} time series.

If not enough measurements were available for age 0.33–3.33, the first 36 monthly recordings were used. Since the performance generally decline somewhat over time, this estimate was adjusted upwards based on the observed median trend for units in the same cohort (similar start years).

4.2 QUANTIFYING TRENDS

As demonstrated in Refs. [1], [4], [5], several different methods exist for calculating trends in wind farm performance, e.g. regression of all CFs against age, average trends for individual units, a capacity-weighted fit of individual trends against

[18] Since SG14 obtained almost identical results with and without LTC, the methodology for estimating mean wind speeds is however not likely to affect their results to any larger degree.

start year, full fixed effects regression and regression of CFs against age for cohorts of units. All methods can in principle be used on both CF_{raw} and CF_{LTC}.

Which of these are most suitable for Swedish conditions and the available data? As already discussed in Section 4.1, LTC of the measurements is, in our opinion, necessary for obtaining trustworthy results for Sweden. Methods not employing LTC were thus discarded.

Quite remarkably, SG14 obtained almost identical trends with regression of all CFs against age as with the other methods. This implies that the CFs of wind farms in the UK (of a certain age) have not increased considerably over time; since the data for newer farms are dominated by low-age observations, the estimated trend would be more negative with this than the other methods had the CFs increased. Modern Swedish wind farms, on the other hand, have considerably higher CFs than older ditto, see Figure 9. Regression of all CFs against age is therefore not relevant for Swedish conditions.

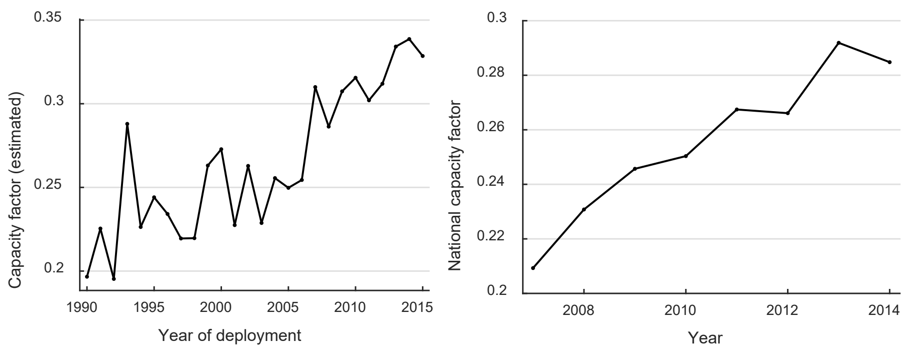


Figure 9. Historical capacity factors (CFs) of wind farms in Sweden; mean estimated CFs (before farm construction) depending on year of deployment (left panel); national, long-term corrected, CF (right panel). Data from Ref. [41].

4.2.1 Linear regression

In this section, linear regression is described relatively briefly and informally. The exposition focus mostly on regression diagnostics and is based on Refs. [42], [43]. The statistical model for a straight line is given by

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

where Y is the dependent variable (CF_{LTC} in our case), X is the independent variable (age), β_0 and β_1 are the intercept and slope and ε is a Gaussian random variable with zero mean and variance σ^2. Given a set of X and Y observations, β_0 and β_1 can be estimated by minimising the residual sum of squares (RSS)

$$RSS = \sum_{i=1}^{N} (y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i)^2 \quad (2)$$

where x_i and y_i are observations from X and Y and the hat symbols over the β's denote that these are estimates of the true (but unknown) parameters. In order to draw conclusions from a straight line fit, several assumptions need to be made [42].

If some of these are severely violated by the data, the linear model might be inappropriate and/or the calculated confidence intervals erroneous.

1) *Existence* – For any value of X, Y is a random variable with finite mean and variance. This is not an issue for our data.

2) *Independence* – The Y-values are statistically independent.

3) *Linearity* – The mean of Y is a straight-line function of X. If this does not hold, patterns in the residuals results.

4) *Homoscedasticity* – The variance of Y has the same expected value independent of the value of X.

5) *Normal distribution* – For any fixed value of X, Y has a normal distribution.

Point 4 and 5 can equivalently be interpreted as normally distributed and homoscedastic residuals. The residuals were analysed in order to answer whether our data reasonably well fulfil these requirements. First, the residuals were plotted against the predicted value. For illustration, some synthetic examples are given in Figure 10. These plots give information on the appropriateness of the linear model and can be used to detect heteroscedasticity of the residuals.

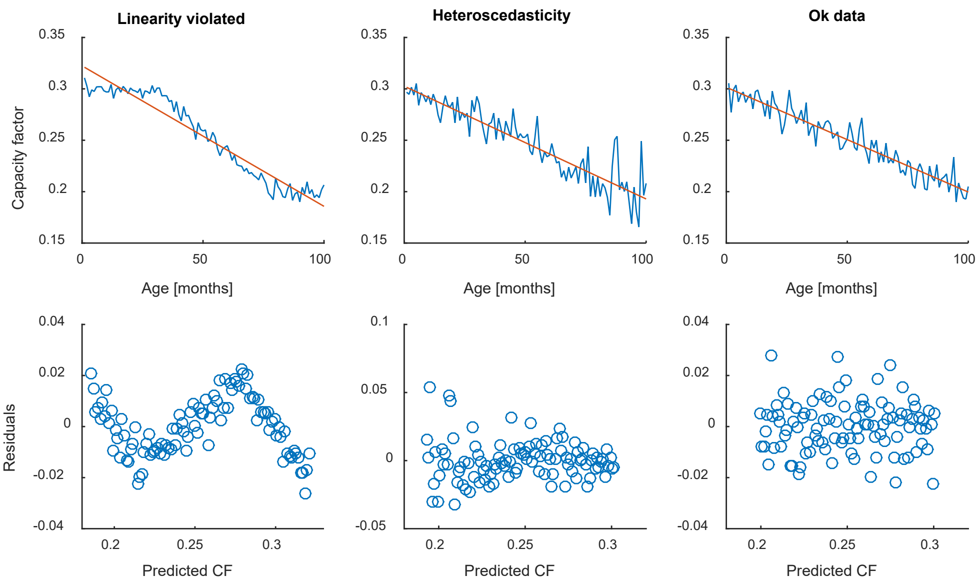


Figure 10. Synthetic examples of capacity factor (CF) time series and linear fits (upper row) and the corresponding residual vs predicted CF (lower row). In all cases, the CF decreases from 0.3 at month 1 to 0.2 at month 100 (plus some random noise). In the leftmost panels, the deterioration is not linear and consequently the linear trend is not appropriate for quantifying the performance decline. In the middle panels, the variance is larger at the end of the measurement period. The trend is not systematically affected, but confidence intervals cannot be properly determined.

Second, the autocorrelation function (ACF) was studied. The ACF is the correlation of a time series with itself, shifted in time with different lags. As an example, the residuals from fitting a straight line to CF_{raw} will almost certainly have relatively strong autocorrelations since there is a seasonality in wind power. With CF_{LTC}, this pattern is ideally removed. As already mentioned, seasonal biases were however present for some units also after LTC, which could be e.g. due to seasonal patterns in wind shear, air density, icing losses and maintenance. As the fictive example in Figure 11 illustrates, neglecting the seasonality may produce erroneous trend estimates. See also Appendix Section 4.2 for an example with real data.

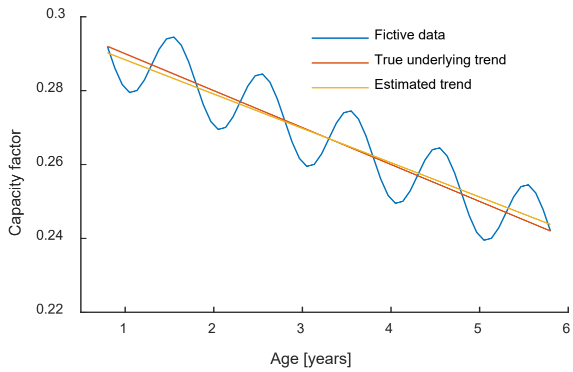


Figure 11. Fictive example demonstrating that if a seasonal pattern exists, the straight line trend estimate might be erroneous. By including a seasonal component in the regression, this problem is alleviated (see below).

Our solution was to also consider a straight line model with a seasonal component:

$$CF_{LTC} = \beta_0 + \beta_1 age + \beta_2 \cos(2\pi \cdot age) + \beta_3 \sin(2\pi \cdot age) + \varepsilon \qquad (3)$$

By including both a sine and a cosine term, the magnitude and phase of the seasonal pattern can be represented. For each unit, the model (with or without seasonal component) with lowest biased corrected Akaike's information criteria (AICc) [44] was chosen. Adding components also for harmonics (4π, 6π, ...) can sometimes be motivated [45], but was found unnecessary in our case.

The assumption of normally distributed residuals was evaluated with normal probability plots. Furthermore, the leverage and DFBETA, which are commonly used techniques to detect outliers, were calculated. The leverage indicates the extremeness of an observation in the range of X-values [42]. In our case, high leverage stem from short measurement periods before or after a long period of missing data. The DFBETA measures the effect on the fitted model parameters from removing an observation. In contrast to the leverage, both X and Y values influence the DFBETA.

As shown in Appendix Section 4.2, the normality and independence assumptions were violated for some units. In particular, for units with several consecutive downtime months, the residuals will be autocorrelated and the residuals for the downtime months more negative than a normal distribution would suggest. The main impact from these violations is that confidence intervals for the trend cannot be properly determined. Since such intervals for individual units is not the primary focus of this report, this was not seen as a major issue. Violation of the linearity is more serious, since the linear slope might then not be representative for the true change in performance (see Figure 10). We therefore also, for comparison, calculated "equivalent trends", see next sub-section.

4.2.2 Chosen methods

The methods that will be used in this work are described in the following paragraphs. For all trend calculations, only ages between four months and 20 years were considered.

For each method, analyses were made for Vindstat and Cesar data and with MERRA, ERA-Interim and ConWx as long-term reference, i.e. for six different cases. This provides additional information on the uncertainties of the trend estimates.

Method 1: Individual trends. For each unit, the linear trend of CF_{LTC} was calculated. As mentioned in the previous section, the best (in terms of lowest AICc) of the following two models were chosen for each unit

$$CF_{LTC} = \beta_0 + \beta_1 age + \varepsilon \tag{4}$$

$$CF_{LTC} = \beta_0 + \beta_1 age + \beta_2 \cos(2\pi \cdot age) + \beta_3 \sin(2\pi \cdot age) + \varepsilon \tag{5}$$

Distributions and summary statistics of the individual trends (β_1) are presented both for all units and for different cohorts depending on start year. If nothing else is specified, we use a three years sliding window (start year 1989–1991, 1990–1992 etc.) in order to not get too few units in each cohort.

Method 2: Combined trends for cohorts. For each cohort, i.e. all units with certain start years coming from a certain dataset and long-term corrected with a certain reanalysis, the model was:

$$CF_{LTC} = \beta_0 + \beta_1 age + \sum_{n=2}^{N} \beta_n I[n] + \varepsilon \tag{6}$$

where n is the unit number of the N units in the cohort and $I[n]$ is an indicator (dummy) variable taking the value one for observations for unit n and zero for all other observations. The parameters β_2–β_N thus controls the level of all units except one (since we also have an intercept β_0) and β_1 represents the slope for the whole cohort. The latter is of course the parameter we are primarily interested in.

Method 3: Equivalent trends. When studying regression diagnostics, it was found that Method 1 was suitable for most units. In some cases, the trend was however not truly linear and e.g. a parabolic curve would fit the data better. Applying different model orders for different units would however be hard to interpret, especially since both the lowest and highest ages vary substantially between the units. In other words, a single parameter (the slope) describing how performance change over time was desired for each unit.

An alternative metric was therefore developed. We consider the first 24 observation months as the baseline and the equivalent trend line was forced to have the same average value during these months as the observations.

Furthermore, the average value of all ages should be the same. The conditions that must be satisfied can be expressed as

$$\begin{cases} \sum_{n=1}^{24} CF_{LTC,n} = \sum_{n=1}^{24} \beta_0 + \beta_1 a_n \\ \sum_{n=1}^{N} CF_{LTC,n} = \sum_{n=1}^{N} \beta_0 + \beta_1 a_n \end{cases} \tag{7}$$

where a is age, n is the month index and N is the total number of observations. Rearranging the equation system gives two equations for the two unknown variables so a unique solution exists. A fictive example is given in Figure 12. A more realistic example of when the linear and equivalent trends differ is when the monthly generation is zero or very low for observations with intermediate ages.

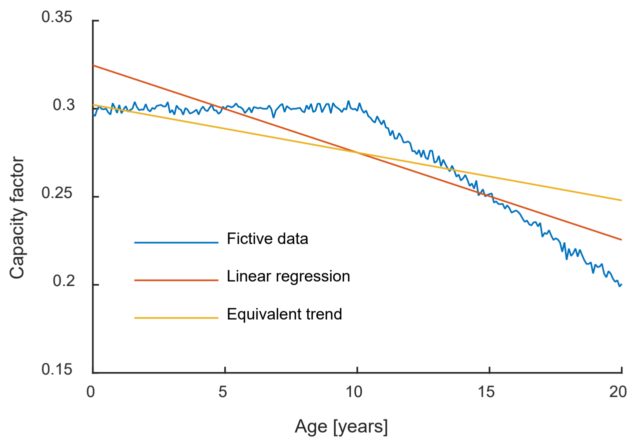


Figure 12. Fictive example demonstrating the difference between linear trend and equivalent trends. In this example, the data clearly suggests that a linear regression is not suitable. The yellow line gives the equivalent trend; a unit with this production would produce an equal amount of energy as the fictive data for the first 24 months as well as for all ages.

4.2.3 Normalisation of capacity factors?

A fundamental choice is whether to study trends in absolute or normalised CFs. For instance, if one farm with CF 0.20 at age 1 has CF 0.15 at age 20, is a farm with CF 0.40 at age 1 more likely to have CF 0.35 or 0.30 at age 20? In order to answer whether trends are more consistent in absolute or relative terms, a test was performed. For each cohort of Vindstat data, the trend of absolute and normalised CFs were computed for each unit. Subsequently, a leave-one-out-cross-validation (LOOCV) method was employed to predict the average CF during the last 30 months based on early age samples of the unit in question and the median trend of the other units in the cohort.

The standard deviations of the prediction errors were very similar when using absolute and normalised CFs, around 0.02 in absolute terms. In other words, the method employed seems to be of little importance for the robustness of the results. A problem with using normalised data is however that some units start reporting as late as age 10–15. How should the CFs be normalised in these cases? If one uses the first months with data, the normalised generation might be overestimated. If one uses some assumption on how much the performance has deteriorated between age 0.33 and the age of the first months with data, the results are influenced heavily by this assumption. We therefore perform all trend analyses with absolute CFs and consider these results as the most robust and important.

For the benefit of the reader, the results were however also transformed to normalised form. Two options in these normalisations are to use the estimated long-term CF for each unit or, as SG14, the average for all units. A disadvantage with the former approach is that some units with low CFs will be very influential for the overall results. The latter approach is more robust but fails to represent the evolution of CFs in Sweden. We take a middle ground and normalise with the cohort average of estimated long-term CFs before deterioration (see Section 4.1). It should be stressed again that these calculations requires assumptions to be made when low-age observations are not available. The normalised trends are thus somewhat more uncertain than their absolute counterparts.

4.2.4 Downtime and efficiency

There is no reliable information on failing components for all units and thus it is not possible to study in detail why performance declines. However, an allocation of a negative trend between increased downtime and worsened efficiency in general (i.e. poorer power curve) is possible. The original plan was to use downtime data from the Vindstat dataset to do this. Internal discussions however revealed that this data is not reliable for all WTs and consequently another approach, based on hourly Cesar data, was taken.

The evolution of downtime over time was studied by analysing the share of zero production in the hourly data. Only periods of zeros which are assumed to be due to actual downtime were considered, i.e. low-wind periods with zero production should not be considered downtime. The methodology for finding such periods was the following:

1) Calculate hourly, fictive output based on ConWx data and the unit characteristics (see Section 4.1).
2) For periods with zero actual production for ≥ 5 hours, calculate the corresponding mean production of the ConWx series.
3) The period is considered as downtime if at least one of the following conditions are fulfilled:
    a. The period is longer than one week
    b. The period is ≥ 12 hours and the corresponding ConWx mean is at least 5% of the rated capacity.
    c. The period is ≥ 5 hours and the corresponding ConWx mean is at least 15% of the rated capacity.

Note that high-wind stops are not properly handled, i.e. these may falsely be classified as downtime due to technical problems. Such stops are however very rare in Sweden.

In general, there is a trade-off between false positive and false negative downtime classifications. If, for example, the threshold of 5 hours is lowered to 3 hours, additional downtime events will be correctly captured at the expense of increasing the number of events falsely classified as downtime. The parameters were trimmed using one year of recorded downtime data for nine WTs, provided by OX2. An example of the results is shown in Figure 13. Relative to the recorded downtime for all WTs, there was 9% false negative and 8% false positive errors with the chosen parameters. Note, however, that only stops that lasted whole hours were considered since it is not possible to detect shorter stops in hourly Cesar data. These short stops consist around 10% of the total downtime for the OX2 WTs. If the occurrence of short stops change with age in the same way as longer stops, the influence from changed downtime on the overall performance trend will thus be somewhat underestimated.

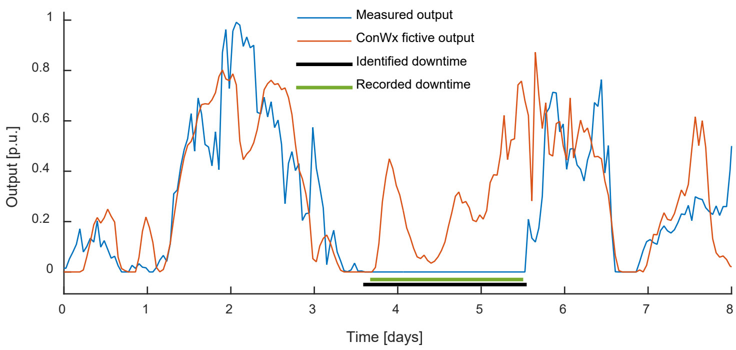


Figure 13. Illustration of method for identifying downtime. During day 3.5–5.5, the measured output was zero but ConWx fictive output was relatively high. Consequently this period was classified as downtime. The zero output period at the end of day 6 was (correctly in this case) not considered downtime since the ConWx output was also very low during these hours.

The methodology was applied to the 94% of all units in the Cesar dataset corresponding to only one WT; for units with more than one WT, it is not possible to identify downtime periods if not all WTs have zero production simultaneously. Linear trends were subsequently calculated for monthly averages of both the original data and for data with the identified downtime periods removed. The difference between these trends is our estimate of the contribution from changes in downtime to the overall trend. Per definition, the remaining trend is explained by changes in the efficiency, e.g. a lowered aerodynamic efficiency if the trend is negative.

4.2.5 Patterns in trends

Although not the primarily objective of this report, it is of course interesting to study whether any clear patterns can be seen in the trends. Preliminary results indicate that older WTs have a faster decline of performance than newer. There are also other factors that can potentially impact the trends:

- Latitude – Wear and tear might be more severe for WTs in the north where icing is a bigger issue.
- Ownership type (private, cooperation, power companies etc.) – Some owners may e.g. have less ambitious maintenance schemes which could impact the deterioration rate.
- WT manufacturer – Possibly, some WTs age better or worse than other.
- Main terrain type – WTs in forest may, for instance, behave differently than WTs close to the coast.
- Mean capacity factor at early ages – Since we study the slope in absolute CFs, the start level might have an impact (a WT with higher CF to start with might lose more output in absolute terms).
- "Tip low" – The distance between ground and lowest blade tip position, i.e. hub height minus rotor radius.
- Longitude, installed capacity and length of the time series were also included.

In total, seven continuous/discrete and three categorical predictor variables (ownership, main terrain type and manufacturer) were thus considered as candidates for explaining the slopes. Each of the categorical predictors were transformed to dummy variables as described in Section 4.2.2, yielding in total ten dummy variables plus the baseline. The baseline was chosen as the most common category for each categorical predictor, see Table 8 on page 56. Separate analyses were made for trends calculated from CF_M, CF_E and CF_C. Since several of the metadata fields were only available in the Vindstat dataset, Cesar units were not studied.

In the analyses, it is important to (at least try to) separate the effects from the different variables. Newer WTs, which generally have less negative slopes, are for example more often installed in forested areas. If the start year was not to be considered, it is likely that one would come to the potentially false conclusion that forest farms deteriorates less than farms deployed on farmland. We thus used different forms of multiple linear regression to estimate the effects of all variables simultaneously. It should be recognised that no hypotheses were formulated so the results must be interpreted with some caution. In particular, one should keep in mind that association does not necessarily imply causality [42].

A few extreme observation potentially get very large influence on the regression results which was not desired. We therefore used winsorisation (see e.g. Ref. [46]) of the slopes. Limits of two standard deviations from the mean were used and we thus replaced slopes smaller than μ-2σ (around -0.77 pp/y) with μ-2σ and slopes larger than μ+2σ (around +0.56 pp/y) with μ+2σ.

Four different multiple linear regression models were set up for explaining differences in slopes:

1) A model with all linear terms.

2) A model with selected (see below) linear terms.

3) Separate models for each predictor variable, only controlling for start year. As an example, the coefficient for latitude was taken from the fitted model $slope = \beta_0 + \beta_1 \cdot start\ year + \beta_2 \cdot latitude$.

4) A model with selected linear and interaction terms (see below).

For choosing predictor variables in method 2 and 4, a forward and backward selection procedure based on AIC was employed. Beginning with a constant model, terms were allowed to enter the model if the AIC was reduced with 0.3 or more. Terms could subsequently be removed if the AIC was not increased with more than 0.1. Suitable threshold values for AIC (i.e. 0.3 and 0.1) were determined by using a training and a test set in the same manner as described in Section 4.3. For the final model, all data were however used.

Interaction terms in method 4 were calculated with Burrill's partial orthogonalisation method [47]. Suppose we have two predictors, U and V. The "raw" interaction term is defined as $U.V = U \cdot V$ (elementwise multiplication). The orthogonalised interaction term UV is subsequently taken as the residuals (ε) from the linear model $U.V = \beta_0 + \beta_1 \cdot U + \beta_2 \cdot V + \varepsilon$. The term UV thus has zero mean and correlates zero with U and V. It was found that orthogonalised interaction terms yielded better models than raw interaction terms; the (adjusted) R squares were higher and the errors/residuals were lower both when evaluating the models with a dedicated dataset and when looking at the fit for all data.

4.2.6 Statistical significance and uncertainty

In order to properly determine confidence bounds for the performance slopes of individual units, several criteria need to be fulfilled as described in Section 4.2.1. Since the trends of individual units is not the main concern of this report, we did not calculate such bounds.

For aggregated results, confidence intervals and statistical significance are however highly interesting. Two examples are:

- Calculate confidence intervals for the mean and median trends of all or subsets of all data.
- Is it statistically significant that the trends of newer units are more positive than for older?

Since the distribution of trends is not normal (extreme trends are more common than they would be had the distribution been normal), we used bootstrap resample methods for calculation of confidence intervals and statistical significance, see e.g. Refs. [48], [49]. A bootstrap sample is a random sample with replacement of equal length as the original data. As an example, (2, 3, 2, 1, 5) could be a bootstrap sample from (1, 2, 3, 4, 5).

Confidence intervals were obtained by drawing 5000 resamples from the original data. The means and medians of these resamples constitutes the empirical bootstrap distributions of these quantities. The 2.5th and 97.5th percentiles of the distributions are taken as the 95% confidence interval. Because of the relatively large datasets, the results are generally similar to those obtained from assuming a normal distribution.

For comparing two different means or medians (see Section 5.2), a similar method was used. Let us for instance consider the median trends for units with start year 2010 (237 units) and start years 2009 or lower (1080 units). The median trends are -0.02 and -0.10 pp/y respectively, i.e. a difference of +0.08 pp/y. 5000 resamples were drawn in which 237 and 1080 trends were randomly picked from all 1317 available. The 5000 resulting differences in the medians of these resamples were computed and the 95% confidence interval of the difference can be estimated as (-0.037, 0.041). The median trend for units with start year 2010 is thus higher than for older units on the 95% confidence level.

It is important to stress that the "statistical uncertainty" is only part of the total uncertainty; given a set of data the statistical confidence interval and/or significance can be calculated, but it is possible that systematic errors exist in the data that makes these metrics invalid. As an example, during some periods, monthly average CFs are not available but rather the average from several months in aggregation. At least for the Vindstat database, such periods are not evenly distributed in time, so before this issue was properly handled, a bias was introduced in the mean and median trends. Another example is that relatively many WTs have recently been rated down to 1500 kW for economic reasons which also introduces an error if not properly accounted for. As described in Section 3.2, we have tried our best to process the data properly and eliminate any sources of bias, but one can of course not be entirely sure that something has not been overlooked. Furthermore, differences exist in the trend estimates from using different datasets and reanalyses for LTC. These differences can be seen as an additional source of information on the uncertainties.

4.3 MACHINE LEARNING

A gradient boosting machine learning technique [50] was used for estimating the rotor diameters and hub heights for units in the Cesar database. For an accessible introduction to machine learning, see Ref. [51] (or Ref. [41] Chapter 3.6 for a five-page introduction). In the following paragraphs, only a few important aspects are discussed.

One advantages with machine learning techniques in general is that non-linear interactions can be captured. In order to not overfit the models, a subset of the data needs to be set aside for determining a suitable model complexity. This validation dataset can also be used for estimating the performance of the model. If the training dataset was to be used for evaluation, overoptimistic results would most likely be achieved.

In order to estimate rotor diameters and hub heights for Cesar units, the installed capacity, deployment year and mean capacity factor were used as predictors. Two-thirds of the Vindstat dataset (randomly sampled) were used for training the models. The remaining samples were set aside for evaluating the model performances. The correlation between predicted and actual rotor diameters was 0.99 and the mean absolute error (MAE) 1.7 metres. For hub heights, the correlation was 0.95 and the MAE 5.0 m. For the purpose of LTC, these errors can be considered small enough. Scatter plots of observed over predicted values in the evaluation dataset are given in Figure 14. After suitable model complexities were determined and the performances evaluated, the final models were trained on all Vindstat data.

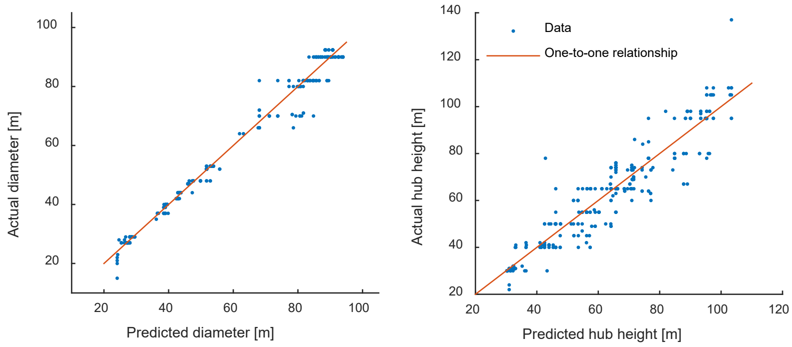


Figure 14. Scatter plot of observed and predicted rotor diameters and hub heights (validation data). A gradient boosting machine learning model with installed capacity, deployment year and mean capacity factor as predictors was used.

A gradient boosting model was also considered for explaining the slope using several variables as predictors (see Section 4.2.5). Since the model did not performed substantially better than the multiple linear regression model with interaction terms and that the results are somewhat harder to interpret, it was decided to not present these results.

4.4 FILTERING

When evaluating the correlation between monthly WIs and actual generation, it might be useful to remove observations corresponding to very low CF_{LTC} since these are likely due to high downtime or other technical problems. The same applies for studies of the variability of CF_{LTC} as a metric for reanalysis performance. One way to do this would be to simply remove observations corresponding to CF_{LTC} below e.g. 50% of the mean value. The problem with this approach is that the performance of the unit might decline with time and thus some observations will be removed although the CF_{LTC} for these months are similar to the neighbouring months. We therefore instead employ a filtering technique which is illustrated in Figure 15.

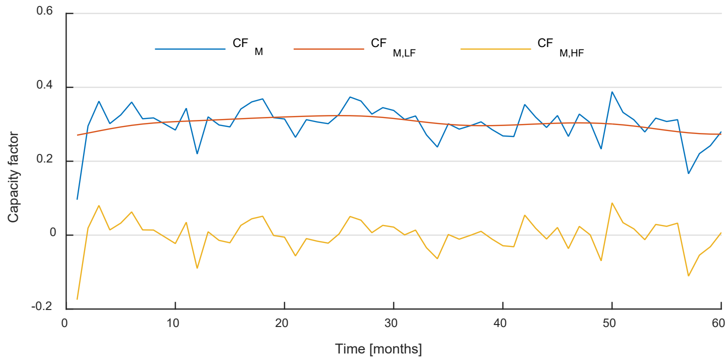


Figure 15. Illustration of filtering for identifying unusually low observations of long-term corrected capacity factors with MERRA as reference (CF_M). First, CF_M is separated into its low- and high-frequency components; $CF_{M,LF}$ and $CF_{M,HF}$ respectively, where $CF_{M,HF} = CF_M - CF_{M,LF}$. An observation is classified as extreme when $CF_{M,HF}$ is below a certain percentage of $CF_{M,LF}$. In this example, at least sample 1 and 57 would be considered unusually low and removed from the analysis.

The low-frequency component was derived from CF_{LTC} by using a windowed-sinc filter applied in both directions. In order to avoid artefacts at the endpoints, the time series were first padded with the mean of the first and last three samples respectively. The added samples were removed after filtering. The high-frequency component was obtained as CF_{LTC} minus the low-frequency component. For an accessible introduction to filters in general, we refer to Smith [52].

4.5 DIFFERENCES AS COMPARED TO SG14

In this section, the main differences of our scope and methods as compared to those in SG14 are summarised.

1) Trends in CF_{raw} were not considered. The reason is that the wind climate has varied considerably over the last years (see Figure 6 on page 30).

2) Observations corresponding to age 0.33–1 years were used in addition to those for age 1–20. The reason is that we have data for individual WTs and that teething issues (lower production during low ages) are not visible in the data after the first few months.

3) Two different datasets and three different reanalyses for long-term correction were used for comparison.

4) In the calculation of monthly WIs, the mean wind speed was determined from the generation data and WT characteristics rather than directly from the reanalyses.

5) As a consequence of the regression diagnostics tests of the seasonal patterns in the residuals of CF_{LTC} in some cases, a seasonal component was added to the linear model when appropriate.

6) A bootstrap method was employed for estimating confidence intervals since the data was not normally distributed. The resulting intervals were however very similar to those obtained with a normal distribution assumption.

7) Partly different methods were used for quantifying the trends. In particular, we did not use regression of all CFs versus age since newer units in Sweden generally have significantly higher CFs than older.

8) Since we had access to hourly data, an allocation of performance decline between increased downtime and worsened power curves was possible.

9) We quantify the association between several variables and the deterioration rate.

5 Results

Six combinations of datasets and reanalyses (i.e., six cases) were considered and several different methods were used for quantifying trends. The body of results is thus substantial and some results were put in Appendix Sections 3–4 in order to facilitate reading the report.

In Section 5.1, reanalyses are compared to measurements. In Section 5.2–5.4, performance slopes are given for the three chosen methods: trends for individual units, trends for cohorts and equivalent trends as described in Section 4.2.2. Section 5.5 displays normalised performance decline. In Section 5.6, an allocation of the deterioration between increased downtime and worsened efficiency is performed. In Section 5.7, finally, systematic differences in trends depending on several variables are investigated.

5.1 REANALYSES

In this section, the three different reanalyses are evaluated by studying correlations between modelled and measured generation. Seasonal patterns and the magnitude of high-frequency fluctuations for CF_{LTC} are presented in Appendix Section 4.1. In conclusion, MERRA and ConWx performs similarly and somewhat better than ERA-I. All three reanalyses were deemed suitable for LTC.

Both hourly data (Cesar 2003–2015) and monthly data (Vindstat and Cesar, 1993–2015) were used for validation. A small number of units outside the coverage of ConWx were not considered. For the analyses of hourly data, ERA-Interim wind speeds were interpolated to hourly resolution using cubic splines. Periods with downtime (see Section 4.2.4), constant production and sudden spikes were not considered. One twelfth of the lowest monthly observations were removed using the filtering method described in Section 4.4.

Results for both hourly and monthly data are given in Figure 16. Correlations for Cesar monthly data (not shown) are very similar to those for Vindstat. Interestingly, the average hourly correlation can be increased with around two percentage points if a weighted combination of the three reanalyses is used. The improvement was however much smaller for monthly data so this option was not utilized. It is important to stress that high correlations does not necessarily imply that the reanalysis is appropriate for LTC. It can, for instance, exist erroneous long-term trends in the data that will produce artificial trends in CF_{LTC} even if the correlations are relatively high. Because of this, it is advantageous to use a few different reanalyses for generating the CF_{LTC} time series and compare the resulting trends.

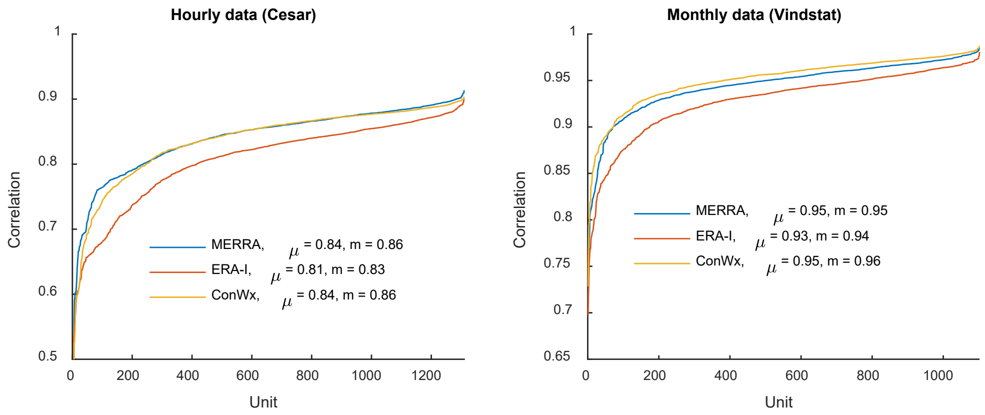


Figure 16. Correlations between hourly and monthly generation calculated from reanalyses and measurements (sorted from lowest to highest). The mean (μ) and median (m) values are given in the legends. Between four and five units have hourly correlations below 0.5, i.e. outside the range shown.

In conclusion, ConWx and MERRA perform roughly equally good and ERA-Interim somewhat worse. Since the horizontal resolution for ConWx is considerably better than for MERRA, one would perhaps expect a little better performance from ConWx. A possible explanation is that we only downloaded 144 ConWx points (i.e., not the closest grid points for all units) and that the WT coordinates are somewhat uncertain. Note, however, that Liléo *et al.* [26] got similar correlations to 42 wind measurements by using MERRA and more high-resolved data. It is thus not obvious that a higher resolution yields better results.

5.2 INDIVIDUAL TRENDS

In this section, trends for individual units are presented. Although the spread is substantial, the means and medians are below zero on the 95% confidence level for all cases. Units with recent start years have less negative slopes, which can mainly be explained by the shorter measurement periods (the slopes are generally less negative during the first years of operation).

The distributions of trends as well as the means and medians for the six different cases are given in Figure 17. Extreme values (outside ±2 pp/y) are not shown. As can be seen, the spread of trends is substantial for each case, but the summary metrics are similar for all cases. Weighting Cesar data on the number of WTs for each unit does not change the means and medians at the precision given in Figure 17.

The average of the six cases is -0.102 pp/y for the mean and -0.101 pp/y for the median. During a 20-year lifetime (age 19.5 versus age 0.5), a unit will thus lose around 1.9 percentage points of CF_{LTC}, e.g. from 0.28 to 0.261. In normalised form, i.e. relative the estimated production at early ages, the corresponding figures are -0.431 pp/y for the mean and -0.402 pp/y for the median. Over 20 years, this implies an energy loss of 4.1% and 3.8% respectively as compared to as-new production.

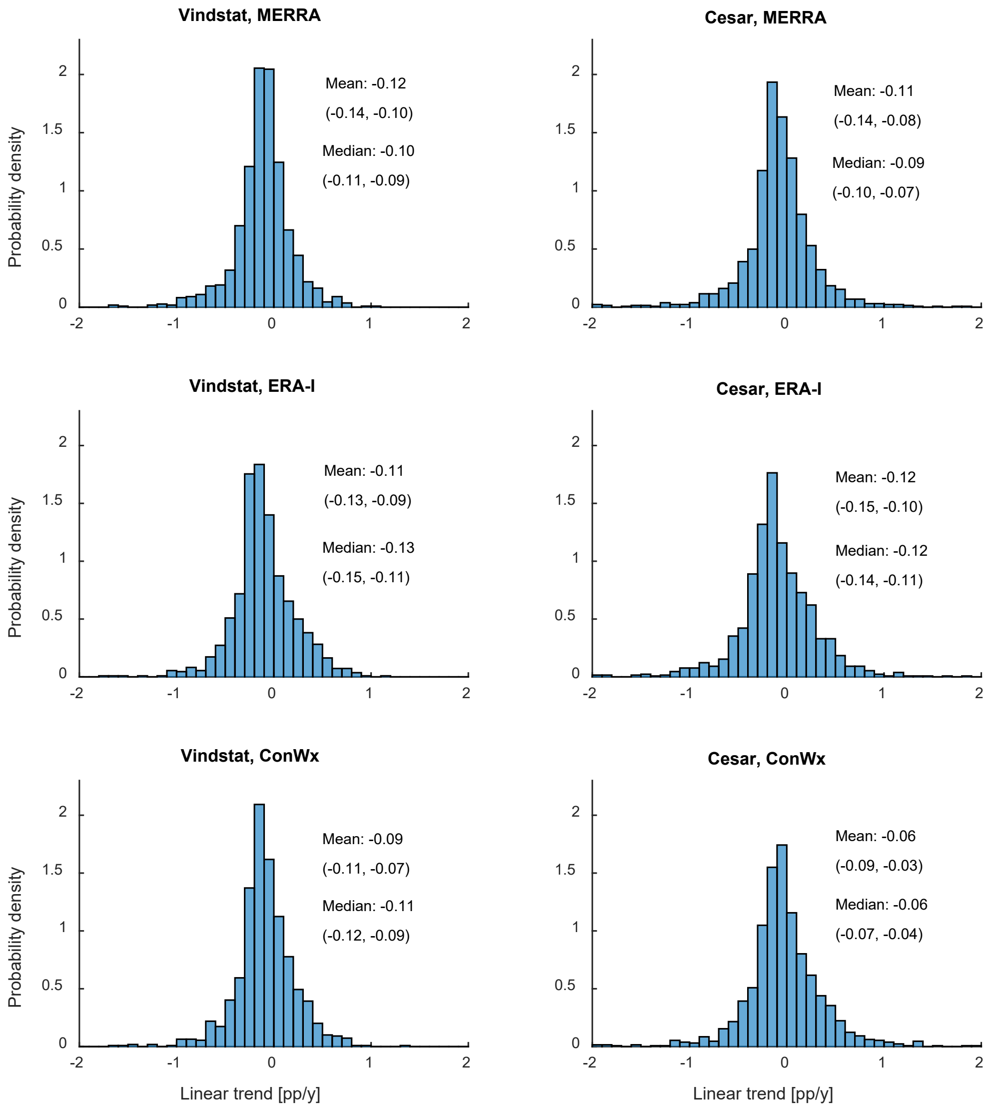


Figure 17. Normalised histograms of all trends for the six different cases. Outliers are not shown (0.7% observations with trend below -2 or above 2 pp/y) but are included in calculations of means and medians. 95% confidence bounds for means and medians are also given.

The 95% confidence intervals for the mean and the median trends were calculated with a bootstrap method since the distributions are not normal. All means and medians are below zero at the 95% level. Tables of means, medians and corresponding confidence intervals for this and other analyses are given in Appendix Section 4.3.

It was observed at an early stage of this project that older units had more negative slopes than those built recently. Figure 18 shows mean and median trends for units in different cohorts for all six cases. The mean trend is often slightly lower than median due to negative extreme trends stemming from e.g. long downtime periods towards the end of the measurements. Figure 19 also shows the 10th, 25th, 75th and 90th percentiles (and has larger ranges of the vertical axes). The spread of trends is larger for the oldest and newest units, which can probably be explained by few units in each cohorts and short measurement periods respectively. The results differ somewhat between the six cases, but are generally in good agreement. For units built around 2005 and earlier, the trends are more negative; the 75th percentiles and sometimes also the 90th percentiles are below zero. For newer units, the trends are less negative; the means and medians (but never the 25th percentiles) are sometimes even positive.

As an average of all six cases, units with start year 2006 and earlier have mean and median trends of -0.18 and -0.15 pp/y, corresponding to a decline of 3.4 and 2.9 CF points over 20 years. In normalised terms, the mean and median trend are -0.76 and -0.65 pp/y, implying a 20-year energy loss of 7.2% and 6.1% respectively as compared to as-new production. The results for the 239 Vindstat units with at least 15 years of data are very similar.

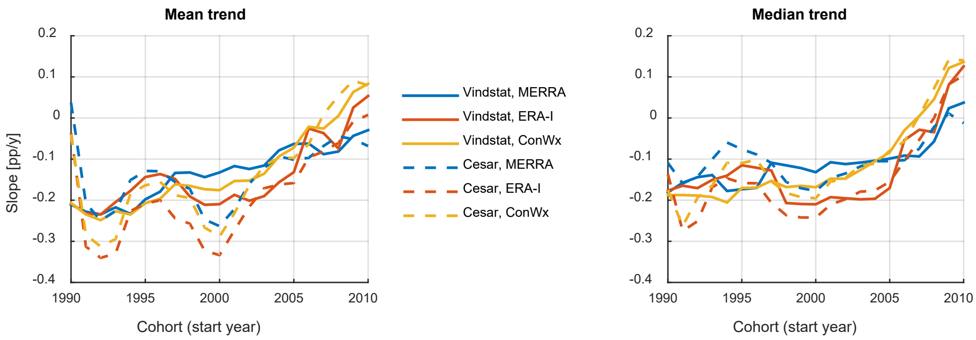


Figure 18. Means and medians of individual trends depending on start year for Vindstat and Cesar and different reanalyses for long-term correction. Results are given for cohorts of a three years sliding window.

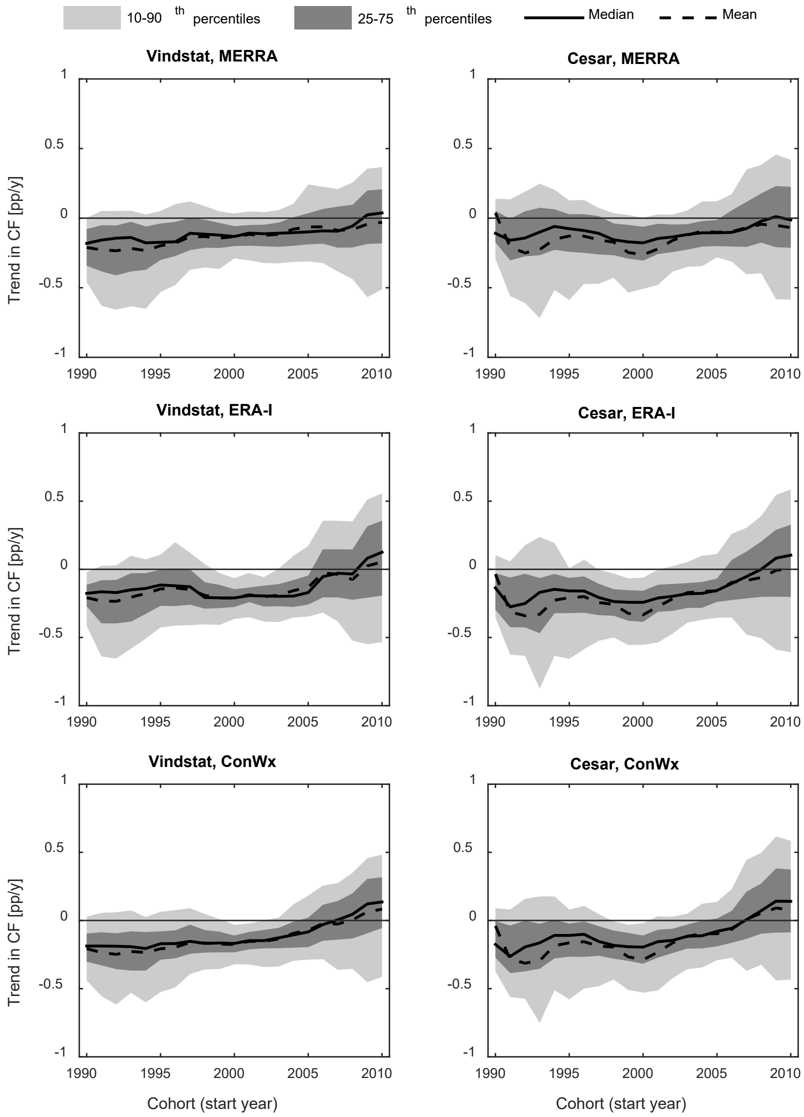


Figure 19. Distribution of individual trends depending on start year for the six different cases. Results are given for cohorts of a three years sliding window.

The question naturally arises whether the slower decline (or even increase) of the performance for units with more recent start years is caused by improvements in technology or by the fact that only data for low ages were available. The full answer to this question can of course only be given 10–15 years from now. A comparison between the trends for newer and older units during the first years of operation can however give some clues.

Trends were calculated for age 0.33 – 5 years for all units with at least 40 months of data during this period. Many older Cesar units were thus excluded since the data collection started in 2003. Figure 20 shows the differences between the mean slopes of units with a certain start year (no sliding window) and all other units. Confidence intervals for rejecting the null hypothesis that no significant difference exist are also given. As before, these intervals were calculated with a bootstrap method. The oldest units in the Vindstat dataset seem to have a steeper decline during the first five years of operation, but otherwise there is no systematic pattern. In particular, newer units does not consistently perform better than others during the first years of operation. The patterns for median trends (not shown) are similar.

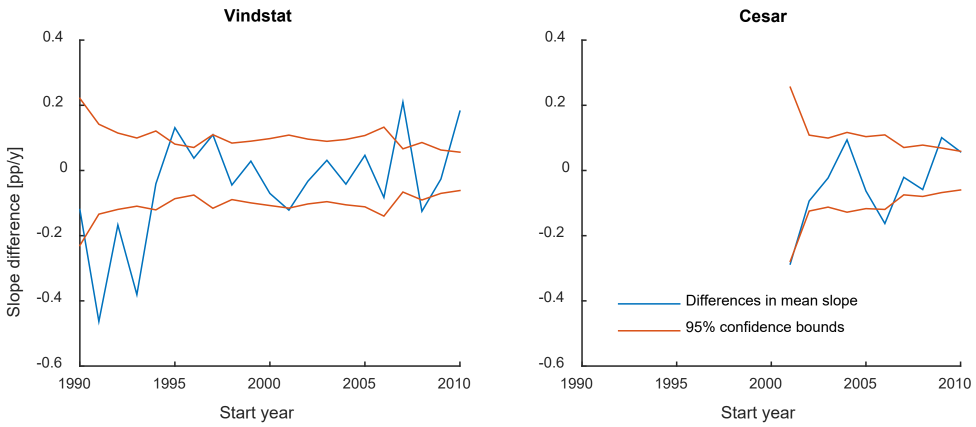


Figure 20. Difference in mean trends for the first five years of operation between units with a certain start year and all other units. A positive difference implies that units with a certain start year have more positive trends than the other units. Confidence bounds for rejecting the null hypothesis that no significant difference exist are also given. The results are given for trends using all three reanalyses (CF_M, CF_E and CF_C).

5.3 TRENDS FOR COHORTS

This and the next section give results for cohort trends and equivalent trends, i.e. methods 2 and 3 as defined in Section 4.2.2. Since the results are very similar to those presented in the previous section, emphasis in the remainder of the report will be on trends for individual units.

Figure 21 shows the resulting slopes from cohort regression for all six combinations of datasets and reanalyses. Weighting on the number of WTs (not shown) has a small impact on the results, considerably smaller than the differences between results for different datasets and reanalyses.

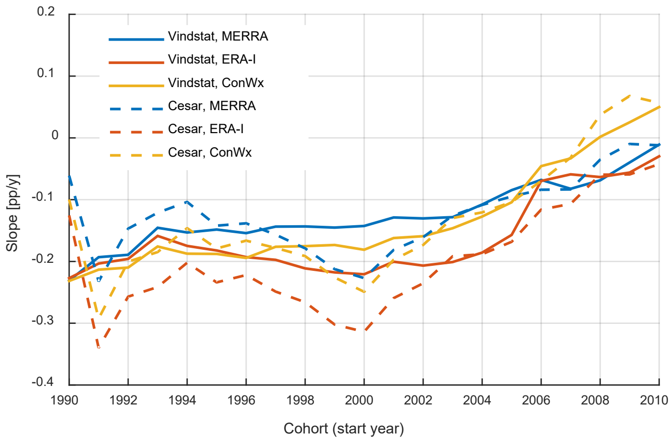


Figure 21. Slope of regression of capacity factors versus age for all units in cohorts.

By visual inspection, the slopes obtained by cohort regression are similar to the mean of the individual trends (Figure 19). The similarity of the different methods will be demonstrated more clearly in the next section.

5.4 EQUIVALENT TRENDS

Equivalent trends can be a useful complement for assessing the robustness of the linear regression methods, in particular for units with poor regression diagnostics. As the results in Table 6 and Figure 22 show, very similar results are obtained with the equivalent trend method and the other two methods.

Table 6 gives mean and median slopes in absolute CF_{LTC} for linear regression and equivalent trends (i.e. Method 1 and 3 as described in Section 4.2.2). Results, both in absolute and normalised form, with confidence intervals are given in Appendix Section 4.3.

Table 6. Comparison of mean and median slopes for linear regression and equivalent trends. All slopes are given in percentage points per year in absolute CFs.

	Mean slope		**Median slope**	
	Lin. reg.	**Eq. trend**	**Lin. reg.**	**Eq. trend**
Vindstat, MERRA	-0.12	-0.11	-0.10	-0.10
Vindstat, ERA-I	-0.11	-0.12	-0.13	-0.13
Vindstat, ConWx	-0.09	-0.10	-0.11	-0.11
Cesar, MERRA	-0.11	-0.09	-0.09	-0.09
Cesar, ERA-I	-0.12	-0.13	-0.12	-0.16
Cesar, ConWx	-0.06	-0.06	-0.06	-0.10
Average	**-0.10**	**-0.10**	**-0.10**	**-0.12**

In Figure 22, the different methods are compared for different cohorts depending on start years. The results are given as averages of all six cases (combinations of datasets and reanalyses). Results from the cohort regressions are given in the left panel since this method has more in common with the mean of individual trends than the median.

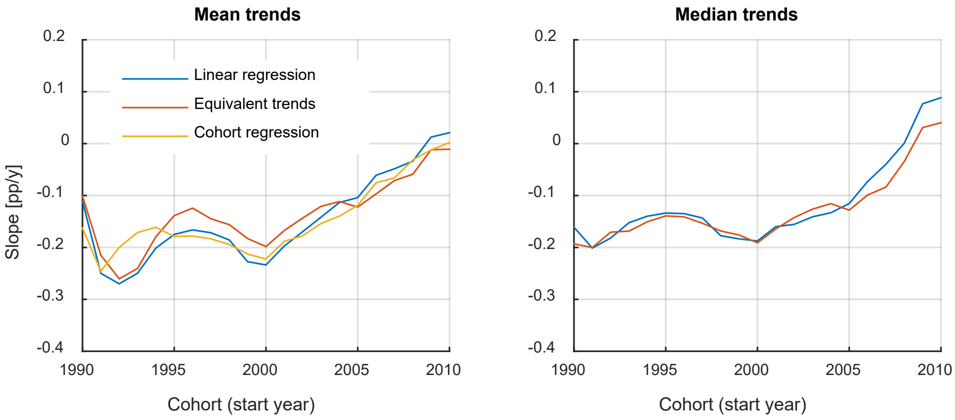


Figure 22. Comparison of mean and median slopes from linear regression and equivalent trends. The average for all six cases (combinations of datasets and reanalyses) are given for different cohorts. For comparison, the average slopes for regression of all data in cohorts (Section 5.3) are also shown in the left panel.

5.5 NORMALISED PERFORMANCE DECLINE

As discussed in Section 4.2.3, there are some caveats with presenting results in normalised form. This is particularly true for Cesar data, since observations for low ages are often not available. For the benefit of the reader, normalised performance declines for all Vindstat units in aggregation are however presented below. The CF_{LTC} time series were normalised to the mean CF_{LTC} of each unit during age 0.33–2.25. A few units with less than twelve monthly low-age observations were removed from the analysis. Figure 23 shows the average “performance factor”, i.e. normalised CF_{LTC}, of all Vindstat units. Note that the number of samples is large for early ages while the results for e.g. age 19 are based solely on observations for units built 1996 or earlier.

The three curves in Figure 23 correspond to an energy loss of 4.4%–5.8% as compared to a constant, unity performance factor. Weighting the results on installed capacity gives around 0.4 pp lower energy loss. The economic loss is smaller, 3.3–4.3% assuming a real interest rate of 6% and no capacity weighting. The results are almost identical if only units built before 2007 are considered, which strengthen the conclusion in Section 5.2 that the performance during the first years is similar for all cohorts. In Appendix Section 4.4, results are presented for Vindstat units with different start years. Results are also given for the Cesar units with low-age observations available for normalisation.

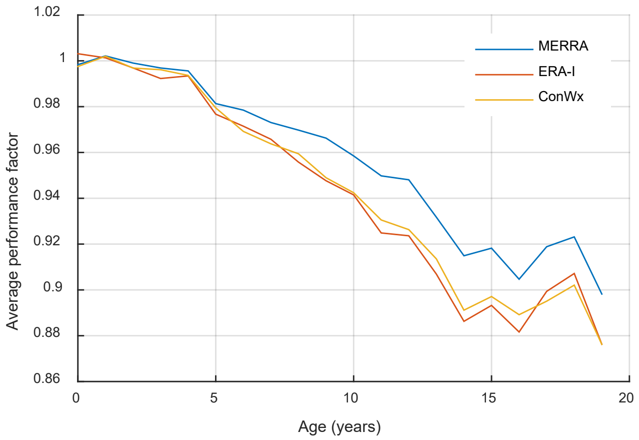


Figure 23. Average performance factors (PF) for all Vindstat units. PF = 1 corresponds to the average capacity factor during age 0.33–2.25.

5.6 DOWNTIME VS EFFICIENCY

In this section, results are given on the evolution of downtime with age. We conclude that downtime increases with age and that this accounts for around 1/3 of the observed performance decline. A description and evaluation of the methodology for detecting downtime can be found in Section 4.2.4. Since hourly data is a necessary input for identifying downtime periods, this section is based solely on Cesar data.

The total downtime was estimated at 4.0%, which seems reasonable in comparison to earlier studies (see Section 2.3). Figure 24 shows how downtime change with age, both for all data (bold, blue line) and for four different cohorts comprising WTs with different start years. The downtime generally increases with age; from around 3.2% at age 0–4 years to 5.9% at age 14–19 years (all units). The increase is statistically significant for all cases but the cohort with start years 2006–2010. Since more low than high age observations are available, the 4.0% given above might be an underestimation. The average of the 20 yearly downtime values (bold, blue line in Figure 24) is 4.5%, which is likely a better assessment. The estimated energy loss is somewhat higher; 4.8% as the average of the 20 yearly values.

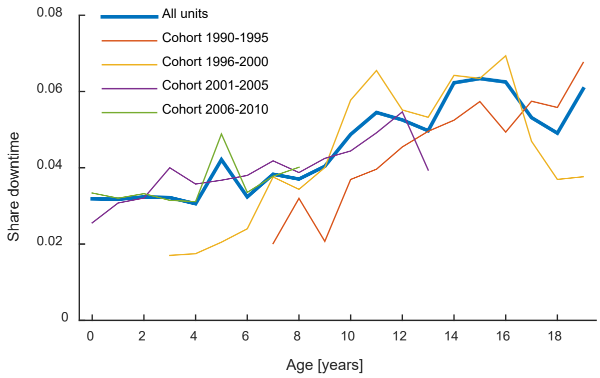


Figure 24. The downtime generally increases with age; from around 3.2% at age 0–4 years to 5.9% at age 14–19 years (all units). Since units with different start years might have different start levels of downtime, downtime vs age is also plotted for cohorts with different start years.

As can be seen in Figure 25, downtime is higher in the winter months, especially for WTs in the north (>60°N). This can most probably be explained by stops due to icing. A more detailed figure is provided in Appendix Section 4.6.

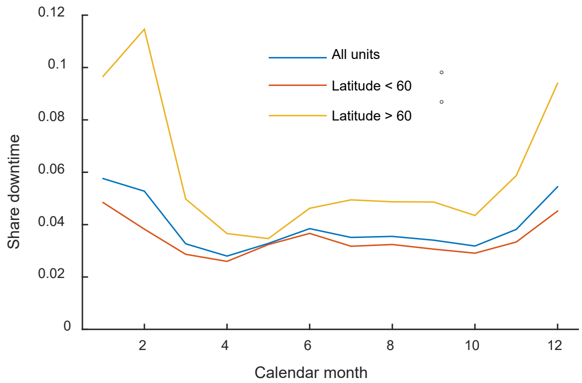


Figure 25. The downtime is generally larger during winter, especially for units in the north. For all data (age 0.33 – 20 years), the estimated downtime is 4.0%.

Since downtime generally increases with age, it should come as no surprise that excluding downtime data gives less negative trends. This is particularly true for units with a steep decline since these declines are primarily caused by increased downtime. Consequently, the mean trend is increased more than the median trend by excluding downtime data, see Table 7 (units with start year 2006 or earlier). As can be seen from the example in Figure 26, the distribution of trends for units with certain start years becomes narrower if downtime data are excluded. This can be interpreted as that large variations exist in the downtime trends, but that the evolution of the efficiency varies less. Some additional results are given in Appendix Section 4.6.

Table 7. Impact on mean and median trends by excluding downtime data (units with start year 2006 or earlier). All results are for Cesar data with units with more than one wind turbine excluded. The trends are given in percentage points per year.

	Mean (downtime included)	Mean (downtime excluded)	Median (downtime included)	Median (downtime excluded)
MERRA	-0.16	-0.09	-0.13	-0.08
ERA-I	-0.23	-0.16	-0.19	-0.15
ConWx	-0.18	-0.11	-0.14	-0.10
Average	**-0.19**	**-0.12**	**-0.15**	**-0.11**

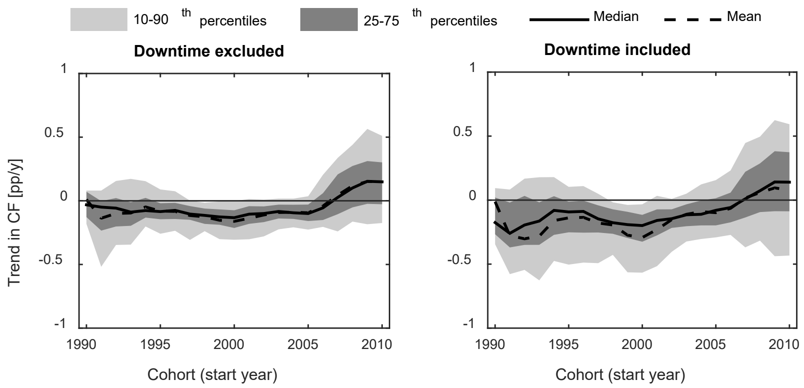


Figure 26. Impact on distribution of individual trends from excluding or including downtime (Cesar data, ConWx for long-time correction). Note that the right panel is not entirely identical to the corresponding panel in Figure 19 since units with more than one wind turbine were excluded here.

It is hard to give a straight answer to how much of the total decline that can be attributed to increased downtime. The median ratios of individual trends excluding and including downtime data are 0.71–0.79 for the three different reanalyses, i.e. one estimate is that downtime accounts for 21–29% of the decline. According to the average results for WTs with start year 2006 or earlier (last row in Table 7), the mean and median trends are 35% and 27% less negative when downtime is excluded. If one sums the energy loss for all units relative their average production during the first 24 months with data, the corresponding figure is 34%. From Figure 27 in the Appendix, the slopes are often around 40% less negative with downtime excluded.

In conclusion, increased downtime accounts for roughly 1/3 of the observed performance decline. For units with strongly negative trends, increased downtime often explains most of the decline.

5.7 INFLUENTIAL VARIABLES

In this section, we investigate whether the slope of individual WTs in the Vindstat dataset can be explained by variables such as start year, manufacturer and main terrain type. A few different multiple linear regression techniques were employed as described in Section 4.2.5. We conclude that start year has a statistically significant (positive) impact on the trends. Some other variables have significant impact in some, but not all, of the analyses performed.

The continuous variables used as regressors were: start year, rated capacity (P), tip low (distance between ground and lowest blade tip position, i.e. hub height minus rotor radius), capacity factor (CF), latitude (lat), longitude (lon) and data length (number of years with measurements). All these variables were normalised, i.e. subtracting the means and dividing by the corresponding standard deviation, for easier interpretation of the results. The categorical variables are listed in Table 8.

Table 8. Categorical variables used for regression of slopes. The baselines are given in italics on the first row and the dummy variables on the following rows. The number *n* of WTs in each category is indicated in parenthesis.

Ownership	Manufacturer	Dominant terrain type
Wind power company (n=467)	*Vestas (n=516)*	*Open landscape (n=644)*
Power company (“power”, n=163)	Enercon (n=306)	Water (sea or lake, n=278)
Other company (“company”, n=111)	WindWorld (“WW”, n=143)	Forest (n=182)
Private (n=191)	Other (n=139)	
Cooperation (“coop”, n=131)		
Other (n=41)		

We begin by giving results for regression of slope on all linear but no interaction terms (Figure 27). Confidence intervals based on a normal distribution assumption, which is reasonably valid in this case, are also indicated. Note that the betas for continuous and dummy variables cannot be interpreted in same way. The normalised start years e.g. range between -1.9 and 1.3, so a beta of 0.072 implies a predicted difference of 0.24 in slope between the oldest and newest WTs. Other continuous variables have other ranges although the means are always zero and the standard deviations always one. The dummy variable betas corresponds to shifts in the level, e.g. around +0.09 for WTs in forest.

Start year has a statistically significant and positive impact on the slope, i.e. newer WTs have less negative slopes as already noted in previous sections. This is true for all three cases (LTC with MERRA, ERA-I and ConWx). WTs in forests have more positive slope than the baseline (open landscape), in average +0.09 pp/y, on the 95% confidence level. Apart from these two variables, no variable has a statistically significant impact for all three cases, although CF and data length are close to meet this criteria. Power companies and other companies seem to perform worse than wind power companies (the baseline). Other manufacturers also perform worse than Vestas. It should be mentioned that this category mainly comprises WTs from former manufacturers such as Bonus, NEG Micon and Danwind, i.e. not contemporary competitors to Vestas and Enercon.

The regression models with all linear terms have adjusted R^2 ranging from 0.10 (MERRA) to 0.18 (ConWx). Most of the differences between the slopes can thus not be explained by the predictors; the stochastic variations are large. In comparison to the confidence intervals for the betas, the impact from not using winsorisation (i.e. that extreme slopes are left unaltered, see Section 4.2.5), is small.

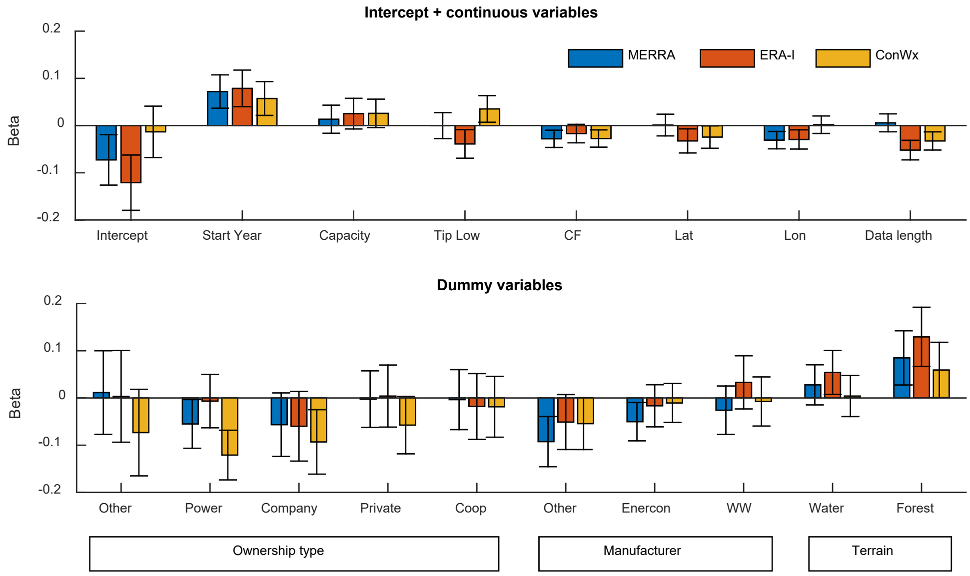


Figure 27. Coefficients in multiple linear regression model (all linear terms but no interactions). See Table 8 and text for variable explanations. Confidence intervals are also indicated.

We also tested to select appropriate predictors out of the ones listed above and to use separate models for all predictors, only controlling for start year (method 2 and 3 as described in Section 4.2.5). The results were relatively similar to those given in Figure 27, see Appendix Section 4.5 for details.

Next, we consider regression models with both linear and interactions terms, chosen with a forward and backward selection procedure. In terms of adjusted R^2, these models perform substantially better than the models without interaction terms (0.27–0.39 as compared to 0.10–0.18). The main effects (betas for linear terms) for the selected variables are presented in Figure 28. The most important differences as compared to Figure 27 are that data length and "other manufacturer" have more negative main effects and that forest has less impact (forest is only selected as predictor for slopes obtained with MERRA data for LTC).

The interaction terms, see Appendix Section 4.5, are generally quite hard to interpret, especially since we have so many predictors. One can however note that the average contributions from these are per definition zero since orthogonalised interaction terms were used.

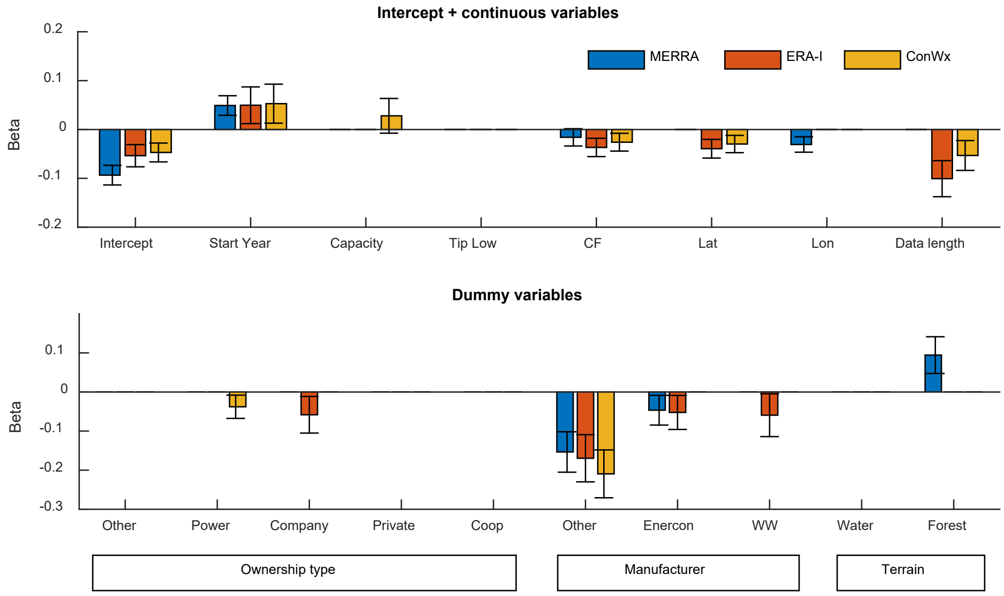


Figure 28. Main effects in regression model with interactions. Predictors were chosen with a forward and backward selection procedure, i.e. not all linear terms are included.

The main conclusion from this section is that there are large variations in the slopes that cannot be explained by the chosen predictors. Some patterns however exist, although only the start year has a significant impact for all possible combinations of regression methods and reanalyses for LTC.

6 Concluding discussion

An important aspect of this work is to be able to give recommendations on suitable performance decline assumptions in energy calculations for future wind farms. For all units, the median decline is around -0.10 pp/y, for units built before 2007, the corresponding figure is -0.15 pp/y and for units built in 2008–2010 we have +0.08 pp/y. Which of these is the best estimate? Since new units do not behave significantly different from older during the first five years of operation, it is reasonable to assume (but of course impossible to truly know) that new units will behave like older when they age, i.e. that -0.15 pp/y is the appropriate figure. Furthermore, it seems that WTs with higher CFs have, everything else equal, somewhat steeper declines (in absolute terms). This is related to the discussion of normalisation; if a WT with CF 0.20 at early age has a CF of 0.15 at age 20 (a reduction of 25%), will a WT with CF 0.40 at early age more likely has CF 0.35 or 0.30 at age 20? The analysis presented in Section 4.2.3 showed that the truth is probably somewhere in between.

Based on the discussion above, our best estimate for current and future farms is a performance decline in the range 0.10–0.20 pp/y over the whole lifetime, including both worsened efficiency and increased downtime. We recommend that a value closer to the higher end should be used for high-CF turbines. When these assumptions are applied to wind energy calculations, early-age downtime losses must be used. If lifetime-average downtime losses was to be added, losses due to increased downtime would be double-counted. Some examples of losses in lifetime production and revenues under different assumptions are given in Table 9.

Table 9. Examples of losses in 20-year lifetime production and revenues for potential new farms given different assumptions. Energy and revenue losses (assuming 6% real interest) are given in relation to those assuming as-new capacity factors (CFs) for the whole lifetime.

As-new CF	Absolute trend	Energy loss	Revenue loss (r = 6%)
0.3	-0.10 pp/y	3.2%	2.5%
0.3	-0.15 pp/y	4.8%	3.8%
0.3	-0.20 pp/y	6.3%	5.1%
0.4	-0.10 pp/y	2.4%	1.9%
0.4	-0.15 pp/y	3.6%	2.9%
0.4	-0.20 pp/y	4.8%	3.8%

Although largely consistent results were obtained with the different methods and datasets/reanalyses, some dissimilarities can be worth discussing. The Cesar and Vindstat databases are partly overlapping, but some units are only present in one of these. In particular, units built after 2003 are not always available in the Vindstat dataset. Vindstat has, on the other hand, production records from 1990 and onwards while measurements are only available from 2003 in Cesar. A comparison between units present in both databases as well as SCADA-data, see Appendix Section 3, showed very small differences in the monthly time series. As an overall judgement, we recommend that equal consideration should be given to the results based on Vindstat and Cesar data. Since the aggregated results, e.g. mean and median trends and patterns of trends versus start year, were similar for the two datasets, this is not a major issue. Regarding the three different reanalyses, the estimated trends were in average slightly less negative when using ConWx data and the differences between trends for different start years were smaller for MERRA. Although the performance (correlation to measurements etc.) were somewhat better for MERRA and ConWx, we recommend that results based on all three reanalyses should be considered. Three reasons for this are that i) the differences in performance were not very large, ii) it is difficult to know whether e.g. a high correlation truly implies good LTC performance and iii) it is intrinsically valuable to study different reanalyses since one gets a better sense of the uncertainties of the results. The three different methods (regression for individual units, cohort regressions and equivalent trends) gave very similar results, which demonstrates the robustness of our estimates.

Based on several different analyses we can conclude that the output during age 0.33–5 years is relatively constant, i.e. only little deterioration occur. A likely explanation is that components have not yet start to deteriorate or fail. Another possible factor is that maintenance contracts with performance guarantees are sometimes signed with the turbine manufacturers for the first years. After the first five years of operation, components begin to fail and, in some cases, less ambitious maintenance schemes are put into operation. Consequently, the performance starts declining.

The deterioration of wind farm performance in the UK is, according to SG14, around -0.43 pp/y in absolute terms, i.e. more than four times the -0.10 pp/y obtained here for Swedish conditions. How can the difference be so large? UK has a harsher climate (stronger extreme wind speeds, more salt spray etc.), which can potentially explain some of the differences. But Sweden, on the other hand, has a colder climate and more icing. Although partly different methods were used here and in SG14, this can hardly explain the dissimilarity. It can be concluded that studies for more countries would be desirable.

In Section 5.7, it was analysed whether variables such as latitude, manufacturer and terrain have a significant impact on the linear slope. Although some patterns were found, most of the variations in slope could not be explained by the predictor variables. There are however many other factors that can contribute to a steep or gentle performance decline:

- Growth and clear-cutting of forests may impact the turbulence intensity and thus the wear and tear.
- Differences in maintenance strategies between different actors.
- Long-term changes in the amount of icing.
- Sector-optimisation of WTs due to noise, shadowing or load may change over time (we only accounted for a general change of WT rating).
- Wake effects can change over time when new WTs are constructed (or dismantled) nearby.
- Bankruptcy of manufacturers makes it hard to get proper service and replace parts. An example is WinWind units, which often have very long downtime periods in recent years.

An important continuation of our work is to identify such factors and improve the strategies for preventing negative performance trends. The results in this report can potentially be used for benchmarking WT performance changes over time.

Acknowledgements

Energimyndigheten, OX2 and Vattenfall are gratefully acknowledged for providing wind power production data. Olle Viotti and Lasse Johansson were initially part of the project on behalf of Sweco, many thanks for your contribution. Jesper Rydén has provided invaluable advice on statistical methods and also commented on parts of the manuscript. The wind power photo on the front cover was taken by Diego Navarro. Finally we would like to thank the reference group for providing insights and comments and to Energiforsk/Energimyndigheten for funding the study.

References

[1] I. Staffell and R. Green, 'How does wind farm performance decline with age?', *Renew. Energy*, vol. 66, pp. 775–786, Jun. 2014.

[2] T. Welte, 'Deterioration and Maintenance Models for Components in Hydropower Plants', PhD Thesis, Norwegian University of Science and Technology, 2008.

[3] R. Kurz and K. Brun, 'Degradation of gas turbine performance in natural gas service', *J. Nat. Gas Sci. Eng.*, vol. 1, no. 3, pp. 95–102, Sep. 2009.

[4] G. Hughes, 'The Performance of Wind Farms in the United Kingdom and Denmark', Renewable Energy Foundation, London, 2012.

[5] 'Paul-Frederik Bach Study'. [Online]. Available: http://www.pfbach.dk/firma_pfb/news_2012_5_e.htm. [Accessed: 03-Dec-2014].

[6] D. Shin and K. Ko, 'Comparative analysis of degradation rates for inland and seaside wind turbines in compliance with the International Electrotechnical Commission standard', *Energy*.

[7] N. Dalili, A. Edrisy, and R. Carriveau, 'A review of surface engineering issues critical to wind turbine performance', *Renew. Sustain. Energy Rev.*, vol. 13, no. 2, pp. 428–438, Feb. 2009.

[8] M. R. Soltani, A. H. Birjandi, and M. Seddighi Moorani, 'Effect of surface contamination on the performance of a section of a wind turbine blade', *Sci. Iran.*, vol. 18, no. 3, pp. 349–357, Jun. 2011.

[9] Y. Zhang, T. Igarashi, and H. Hu, 'Experimental Investigations on the Performance Degradation of a Low-Reynolds-Number Airfoil with Distributed Leading Edge Roughness', in *49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, American Institute of Aeronautics and Astronautics.

[10] M. G. Khalfallah and A. M. Koliub, 'Effect of dust on the performance of wind turbines', *Desalination*, vol. 209, no. 1–3, pp. 209–220, Apr. 2007.

[11] Y. Sinha and J. A. Steel, 'Failure Prognostic Schemes and Database Design of a Software Tool for Efficient Management of Wind Turbine Maintenance', *Wind Eng.*, vol. 39, no. 4, pp. 453–478, May 2015.

[12] A. Stenberg, 'Analys av vindkraftsstatistik i Finland', Master thesis, Aalto University, 2010.

[13] N.-E. Carlstedt, 'Driftuppföljning av Vindkraftverk, Årsrapport 2006', Elforsk rapport 07:14, 2007.

[14] F. Carlsson, E. Eriksson, and M. Dahlberg, 'Damage preventing measures for wind turbines, Phase 1 – Reliability data', Elforsk report 10:68, 2010.

[15] F. Spinato, P. J. Tavner, G. J. W. van Bussel, and E. Koutoulakos, 'Reliability of wind turbine subassemblies', *IET Renew. Power Gener.*, vol. 3, no. 4, pp. 387–401, Dec. 2009.

[16] P. J. Tavner, J. Xiang, and F. Spinato, 'Reliability analysis for wind turbines', *Wind Energy*, vol. 10, no. 1, pp. 1–18, Jan. 2007.

[17] S. Faulstich, B. Hahn, and P. J. Tavner, 'Wind turbine downtime and its importance for offshore deployment', *Wind Energy*, vol. 14, no. 3, pp. 327–337, Apr. 2011.

[18] B. Hahn, M. Durstewitz, and K. Rohrig, 'Reliability of Wind Turbines', in *Wind Energy*, P. D. J. Peinke, P. D.-I. P. Schaumann, and D. S. Barth, Eds. Springer Berlin Heidelberg, 2007, pp. 329–332.

[19] W. Yang, P. J. Tavner, C. J. Crabtree, Y. Feng, and Y. Qiu, 'Wind turbine condition monitoring: technical and commercial challenges', *Wind Energy*, vol. 17, no. 5, pp. 673–693, May 2014.
[20] F. P. García Márquez, A. M. Tobias, J. M. Pinar Pérez, and M. Papaelias, 'Condition monitoring of wind turbines: Techniques and methods', *Renew. Energy*, vol. 46, pp. 169–178, Oct. 2012.
[21] J. Ribrant and L. M. Bertling, 'Survey of Failures in Wind Power Systems With Focus on Swedish Wind Power Plants During 1997 ndash;2005', *IEEE Trans. Energy Convers.*, vol. 22, no. 1, pp. 167–173, Mar. 2007.
[22] K. Harman, R. Walker, and M. Wilkinson, 'Availability trends observed at operational wind farms', in *Proc. European wind energy conference*, Brussels, Belgium, 2008.
[23] N. Conroy, J. P. Deane, and B. P. Ó Gallachóir, 'Wind turbine availability: Should it be time or energy based? – A case study in Ireland', *Renew. Energy*, vol. 36, no. 11, pp. 2967–2971, Nov. 2011.
[24] M. M. Luengo and A. Kolios, 'Failure Mode Identification and End of Life Scenarios of Offshore Wind Turbines: A Review', *Energies*, vol. 8, no. 8, pp. 8339–8354, Aug. 2015.
[25] E. Sharp, P. Dodds, M. Barrett, and C. Spataru, 'Evaluating the accuracy of CFSR reanalysis hourly wind speed forecasts for the UK, using in situ measurements and geographical information', *Renew. Energy*, vol. 77, pp. 527–538, May 2015.
[26] S. Liléo, E. Berge, O. Undheim, R. Klinkert, and R. E. Bredesen, 'Long-term correction of wind measurements - State-of-the-art, guidelines and future work', Elforsk, 13:18, Jan. 2013.
[27] D. Carvalho, A. Rocha, M. Gómez-Gesteira, and C. Silva Santos, 'Comparison of reanalyzed, analyzed, satellite-retrieved and NWP modelled winds with buoy data along the Iberian Peninsula coast', *Remote Sens. Environ.*, vol. 152, pp. 480–492, Sep. 2014.
[28] S. Liléo and O. Petrik, 'Investigation on the use of NCEP/NCAR, MERRA and NCEP/CFSR reanalysis data in wind resource analysis', in *Proceedings for EWEA*, Brussels, Belgium, 2011.
[29] J. Sonderegger, 'Performance of long-term correction at wind power development', Master thesis, Chalmers University, Gothenburg, Sweden, 2011.
[30] M. Brower, M. Barton, L. Lledó, and J. Dubois, 'A study of wind speed variability using global reanalysis data', AWS Truepower, 2013.
[31] W. Skamarock *et al.*, 'A description of the Advanced Research WRF Version 3', 2008.
[32] J. Salmon and P. Taylor, 'Errors and uncertainties associated with missing wind data and short records', *Wind Energy*, vol. 17, no. 7, pp. 1111–1118, Jul. 2014.
[33] M. M. Rienecker *et al.*, 'MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications', *J. Clim.*, vol. 24, no. 14, pp. 3624–3648, Jul. 2011.
[34] D. P. Dee *et al.*, 'The ERA-Interim reanalysis: configuration and performance of the data assimilation system', *Q. J. R. Meteorol. Soc.*, vol. 137, no. 656, pp. 553–597, 2011.
[35] S. Saha *et al.*, 'The NCEP Climate Forecast System Reanalysis', *Bull. Am. Meteorol. Soc.*, vol. 91, no. 8, pp. 1015–1057, Apr. 2010.
[36] S. Kobayashi *et al.*, 'The JRA-55 Reanalysis: General Specifications and Basic Characteristics', *J. Meteorol. Soc. Jpn. Ser II*, vol. 93, no. 1, pp. 5–48, 2015.
[37] I. Staffell and S. Pfenninger, 'Using bias-corrected reanalysis to simulate current and future wind power output', *Energy*, vol. 114, pp. 1224–1239, Nov. 2016.

[38] S. J. Watson, P. Kritharas, and G. J. Hodgson, 'Wind speed variability across the UK between 1957 and 2011', *Wind Energy*, vol. 18, no. 1, pp. 21–42, Jan. 2015.

[39] M. L. Thøgersen, M. Motta, T. Sørensen, and P. Nielsen, 'Measure-Correlate-Predict Methods: Case Studies and Software Implementation', in *EWEC 2007 Scientific Proceedings*, 2007.

[40] J. Olauson and M. Bergkvist, 'Modelling the Swedish wind power production using MERRA reanalysis data', *Renew. Energy*, vol. 76, pp. 717–725, Apr. 2015.

[41] J. Olauson, 'Modelling Wind Power for Grid Integration Studies', PhD Thesis, Uppsala University, 2016.

[42] D. G. Kleinbaum, L. L. Kupper, and K. E. Muller, *Applied regression analysis and other multivariable methods*, Second edition. Boston, Mass.: PWS-Kent Pub. Co., 1988.

[43] T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, Second edition. New York: Springer, 2001.

[44] C. M. Hurvich and C.-L. Tsai, 'Regression and time series model selection in small samples', *Biometrika*, vol. 76, no. 2, pp. 297–307, Jun. 1989.

[45] R. E. Chandler and E. M. Scott, Eds., *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*. Chichester, UK: John Wiley & Sons, Ltd, 2011.

[46] J. W. Tukey and D. H. McLaughlin, 'Less Vulnerable Confidence and Significance Procedures for Location Based on a Single Sample: Trimming/Winsorization 1', *Sankhyā Indian J. Stat. Ser. 1961-2002*, vol. 25, no. 3, pp. 331–352, 1963.

[47] D. Burrill, 'Modeling and interpreting interactions in multiple regression', https://www.minitab.com/uploadedFiles/Content/News/Published_Articles/interactions_in_multiple_regression.pdf, Accessed: 2017-02-14.

[48] B. Efron and R. Tibshirani, 'Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy', *Stat. Sci.*, vol. 1, no. 1, pp. 54–75, Feb. 1986.

[49] I. Rychlik and J. Rydén, *Probability and Risk Analysis: An Introduction for Engineers*. Springer Science & Business Media, 2006.

[50] J. H. Friedman, 'Greedy Function Approximation: A Gradient Boosting Machine', *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001.

[51] R. A. Berk, *Statistical learning from a regression perspective*, 1. Ed. New York, N.Y: Springer, 2008.

[52] S. Smith, *The scientist and engineer's guide to digital signal processing*, 2nd edition. San Diego, CA: California Technical Publishing, 1999.

Contents

1 Supplementary methods

1.1 THRESHOLD FOR MINIMUM MEASUREMENT LENGTH

In order to determine an appropriate threshold for required measurement length, the following test was performed. For all WTs with at least ten years of data, the linear trend was calculated. Trends were also computed for subsets of different lengths and the differences to the trend for all data were stored. Assuming that the trends for the 10+ year time series are correct estimates, these differences represent errors due to short measurement periods. Figure 1 shows an example of the trends for all data and 36 month subsets.

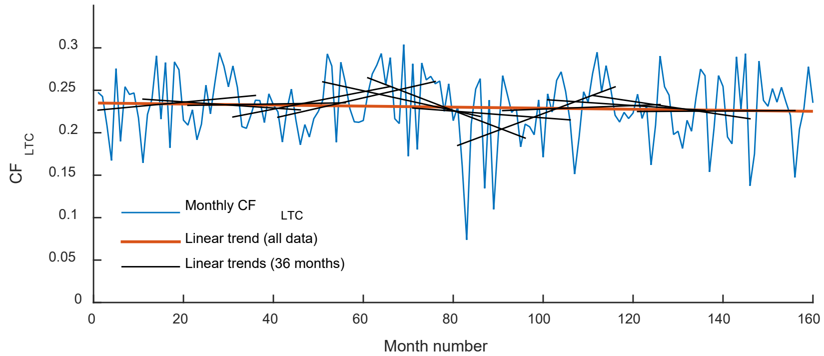


Figure 1. Representative example of trend lines calculated for 36 months subsets and trend line for all data.

For a fictive cohort of WTs, differences were randomly sampled and the error in the median trend estimate was calculated. More precisely, the errors were determined by randomly sampling *N* differences between "true" trends (trend for the 10+ year time series) and trends for subsets of length *L*. This procedure was repeated 1000 times for each combination of *N* and *L*. Figure 2 shows how the error depends on the measurement length and cohort size. Considering the size of our cohorts and the magnitude of the trends we wish to detect, it was decided to use a threshold of 60 months (five years), i.e. the same as in SG14.

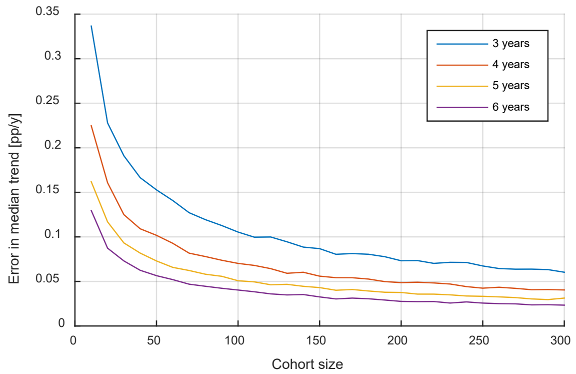


Figure 2. Standard deviation of errors in median trends depending on cohort size (N) and measurement period (L).

1.2 TRANSFORMATION OF POWER CURVES

The ratings of some WTs were changed during the measurement period. Since it was desired to remove the effects of these changes from the trend calculations, a transformation of the time series were made, i.e. the hourly data during the period with higher rating were lowered.

A naïve transformation would be to simply put a cap at the lower rating, e.g. if the rating was changed from 2000 kW to 1500 kW, all values above 1500 kW are changed to 1500 kW. Since the power curve must be smooth before cut-in wind speed, we instead calculated power curves for both ratings and found the power-to-power transform, see Figure 3b. Although our approach is more realistic, the energy difference compared to the naïve transformation is only around 1%.

See also Section 2.1.4 below for an example of an altered time series.

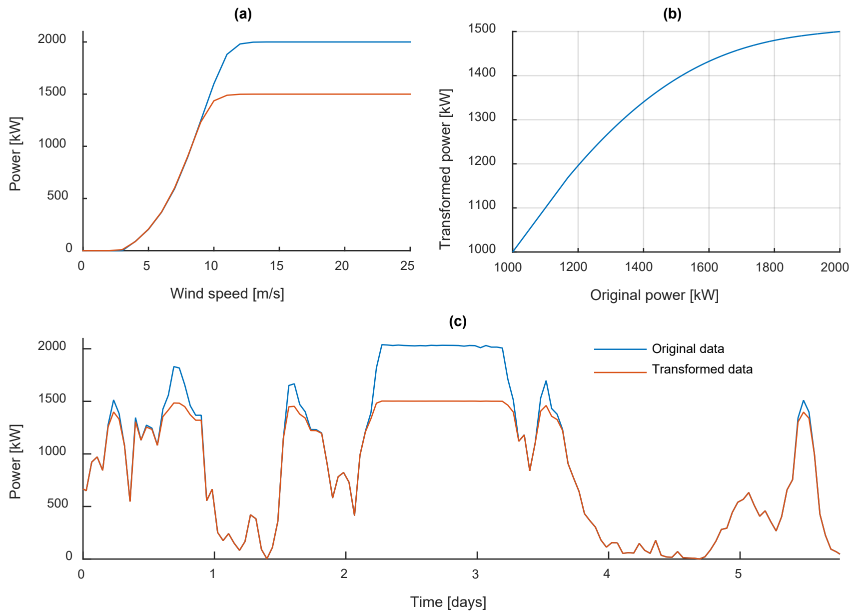


Figure 3. Example of transformation of data from rating 2000 kW to 1500 kW. (a) Power curves; (b) Relation between original and transformed power; (c) Example of hourly time series.

2 Supplementary data treatment

2.1 EXAMPLES OF CHANGES OF DATA

In this section, examples of altered Cesar data, both installed capacity and hourly generation, are given. Some of the changes were also possible to do for monthly Vindstat data.

2.1.1 Change of rating to a constant value

For some units, the ratings in the metadata were simply not correct. Based on a visual inspection, these ratings were changed.

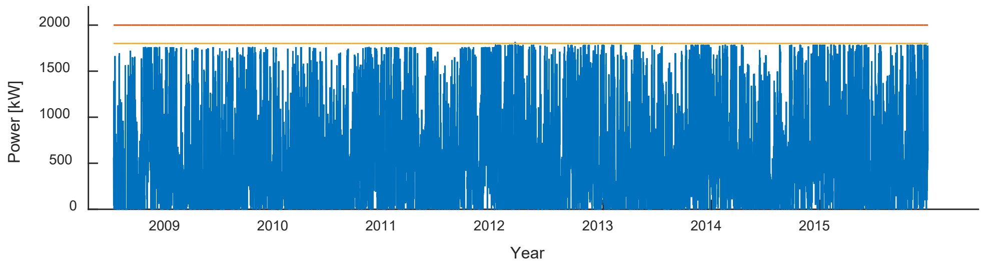


Figure 4. Example of change of rating from 2000 kW (red) to 1800 kW (yellow).

In a few cases, the metadata suggested that the rating had changed, but based on hourly data, this was not the case.

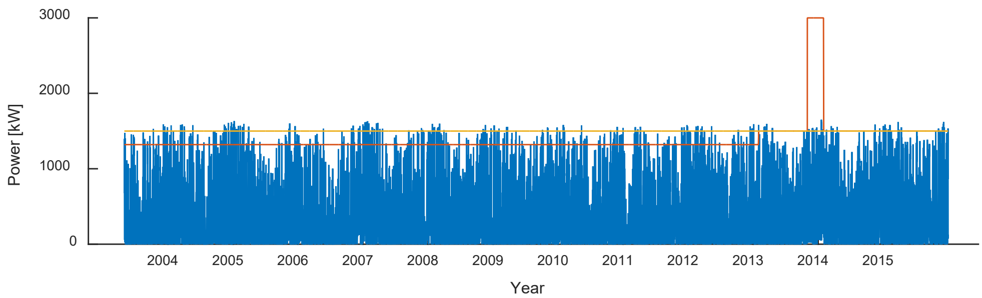


Figure 5. Example of incorrect changes of rating given in the metadata (red line). Here, the installed capacity was changed to a constant value (yellow).

2.1.2 Split time series

It is relatively common that the number of WTs connected to a Cesar unit changed during the measurement period. Often, only one of the segments were longer than 60 months and the short periods were then simply removed. In a few instances, the time series was however split in two.

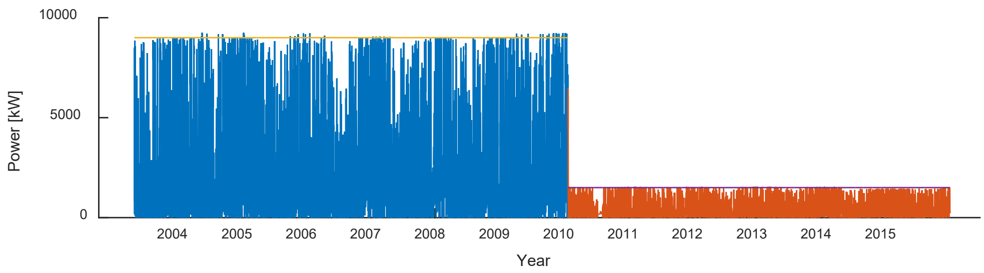


Figure 6. The number of wind turbines changed from six to one around year 2010. The time series was thus split into two segment which were handled separately in the trend analyses.

2.1.3 Temporary down-rating (not to 1500 kW)

Especially for older units, the maximum output can be lower than the rating for extended time periods. If the lower maximum output was not 1500 kW (see next sub-section), we interpret this as down-rating due to technical issues and we did not change the rating.

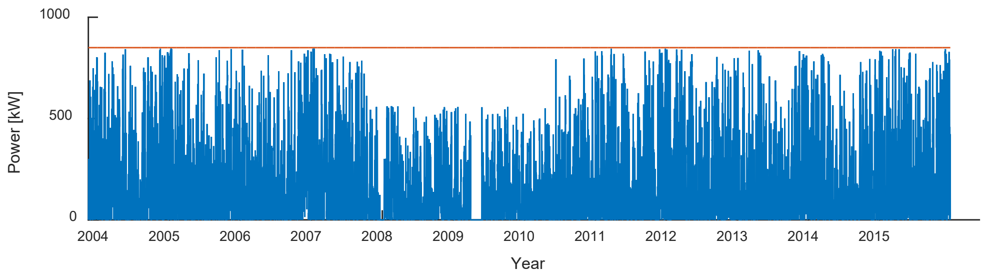


Figure 7. The unit has a lower maximum output for an extended time period. If the lower maximum was not 1500 kW (for each wind turbine), no changes were performed.

2.1.4 The question of 1500 kW

Due to taxes, it might be advantageous to limit the rating to 1500 kW, especially when the electricity and certificate prices are low. We wished to exclude this effect from the WT performance and thus made the transformation described in Section 1.2 in this appendix.

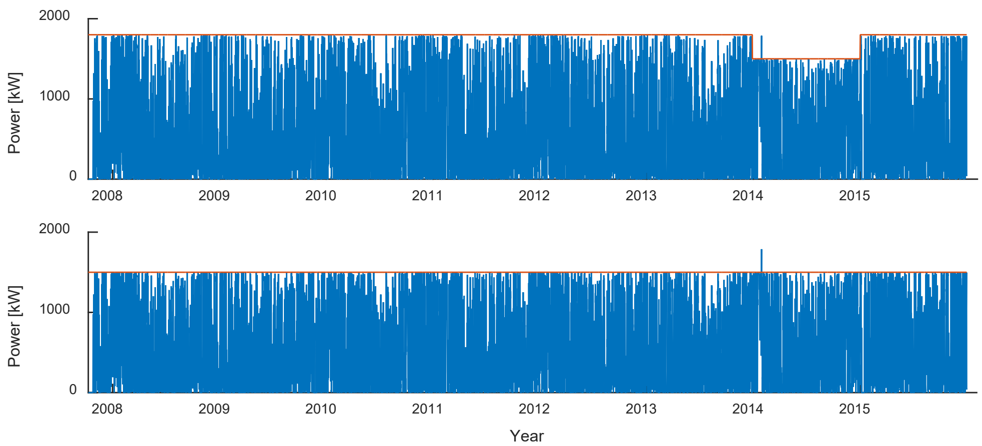


Figure 8. The rating was lowered to 1500 kW in year 2014 (upper panel). All data corresponding to rating 1800 kW was thus transformed based on the method given in Section 1.2 in this appendix (lower panel).

2.1.5 Fictive plants

When problems occur in the reporting system, the generation can be reported later using a "fictive plant". Sometimes, correct hourly measurements are delivered afterwards, but often the generation is reported as a lump sum for only one hour or as a constant production during the period (see Figure 9). In both cases, the total energy is correct, but it is not possible to use the data for analyses of hourly data, e.g. identifying downtime periods. Care must also be taken when measurements are long-term corrected, see discussion in Section 3.2 in the main document.

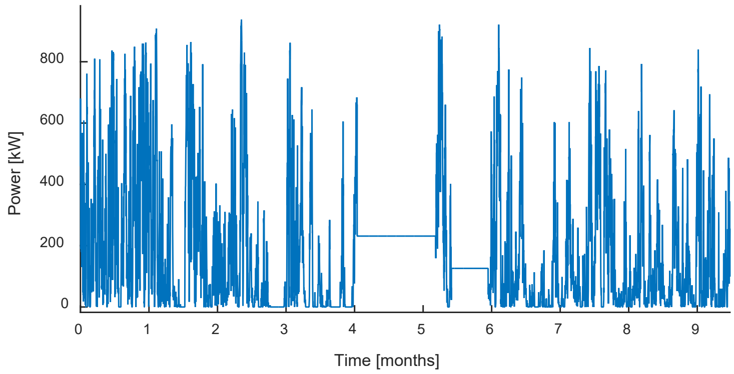


Figure 9. Example of constant production reported from "fictive" plants. If the constant period extends over several months, this has to be accounted for in the long-term correction.

2.1.6 Remove data

For some (38 out of around 1300) units, strange and unexplainable patterns were present in the time series. Such data were simply removed.

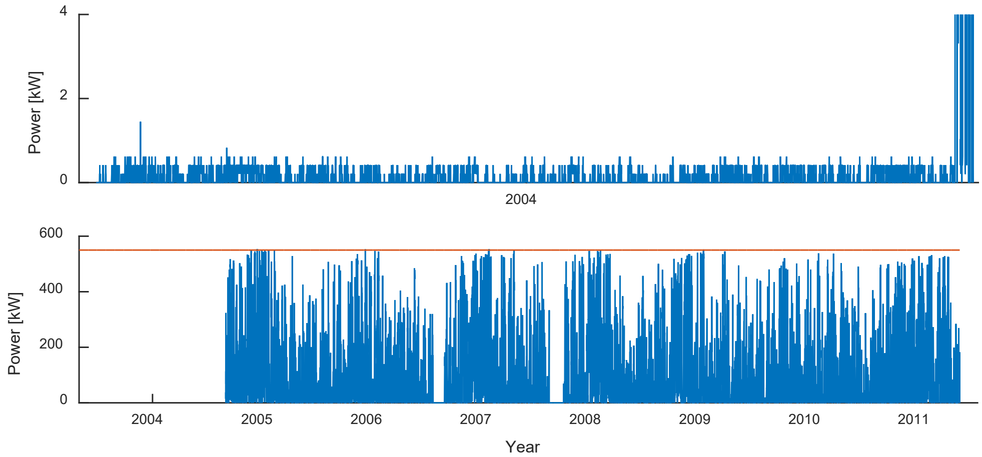


Figure 10. Some strange measurements were removed. In this example, measurements for the whole first year were below 1 kW (upper panel, only the first year is shown). This data were simply removed (lower panel, the whole time period is shown).

3 Comparison between Cesar, Vindstat and SCADA data

In order to validate the quality of the metadata and generation time series, two tests were performed. Firstly, monthly time series and corresponding trends for 97 WTs in three Vattenfall farms were compared to Vindstat and Cesar data. Secondly, 20 random units (present in both Cesar and Vindstat) were compared.

SCADA data (31 months of 10-min measurements) from Vattenfall were available for individual WTs in three different farms. For farm 2 and 3, the corresponding Vindstat time series were shorter than 60 months and thus not analysed in the main report. Here, we however studied trends in raw CFs for the around 55 months with data. The coupling of individual WTs in the three datasets was made based on the monthly time series. Table 1 shows average correlations between the monthly observations from the three different datasets. Overall, the correlations are high, but for farm 2, there are some differences between Vindstat and Cesar when the 55 month time series are compared (first column).

Table 1. Average turbine-wise correlations between monthly observations from the three different datasets.

	Cesar/Vindstat	Cesar/Vattenfall	Vindstat/Vattenfall
Farm 1	0.99	1.00	1.00
Farm 2	0.95	0.99	1.00
Farm 3	1.00	0.98	0.98

In Figure 11-13, linear trends for the three farms are shown, sorted from lowest to highest. Note that the trends can differ substantially between the left and right panels since different time periods were studied and no LTC was performed. Again, the three datasets agree reasonably well, but some differences exist. Apart from potential errors in the measurements, one factor that can explain the differences is the different temporal resolutions. If some measurements are missing in Cesar and SCADA, these were not taken into account in the calculation of monthly CFs. For Vindstat, however, the monthly CFs are based solely on the total monthly generation divided by installed capacity and number of hours each month.

It can finally be noted that the installed capacities, coordinates and start dates were very similar for Cesar and Vindstat metadata.

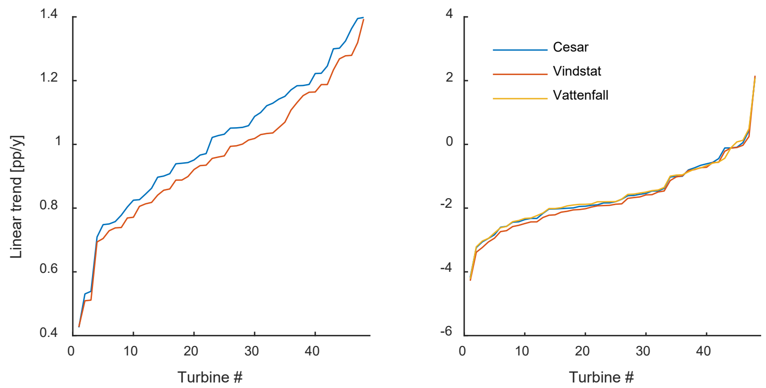


Figure 11. Linear trends for wind turbines in farm 1, sorted from lowest to highest. The left panel shows trends for the around 98 months with concurrent Cesar and Vindstat data and the right panel shows trends for the 31 months with Vattenfall SCADA data. For the latter, the mean trends are -1.6, -1.7 and -1.6 pp/y for Cesar, Vindstat and Vattenfall, respectively.

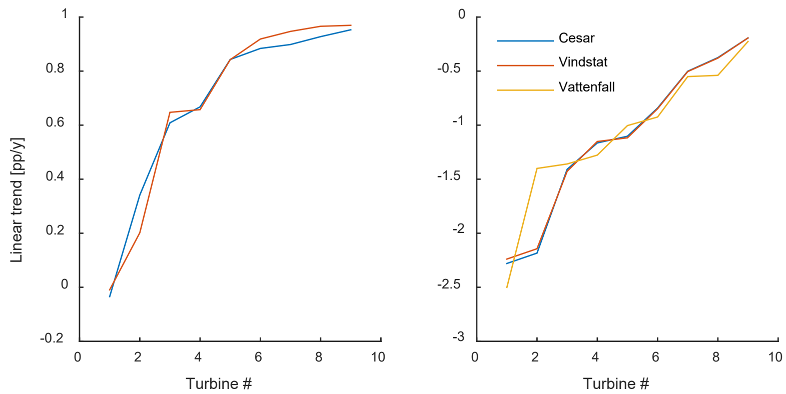


Figure 12. Linear trends for wind turbines in farm 2, sorted from lowest to highest. The left panel shows trends for the around 54 months with concurrent Cesar and Vindstat data and the right panel shows trends for the 31 months with Vattenfall SCADA data. For the latter, the mean trends are all -1.1 pp/y.

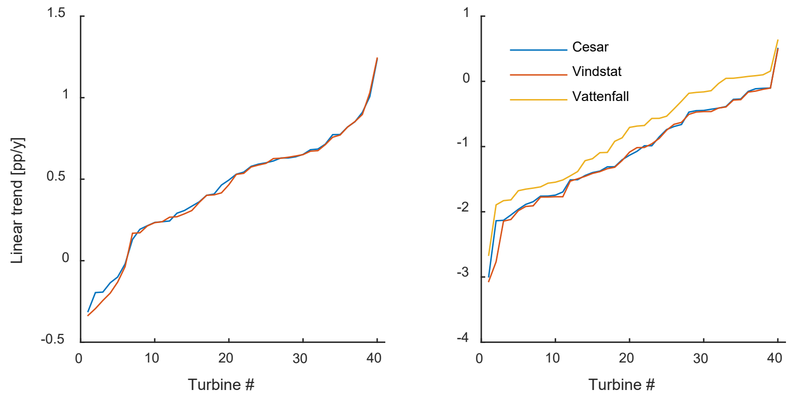


Figure 13. Linear trends for wind turbines in farm 3, sorted from lowest to highest. The left panel shows trends for the 54 months with concurrent Cesar and Vindstat data and the right panel shows trends for the 31 months with Vattenfall SCADA data. For the latter, the mean trends are -1.1, -1.1 and -0.81 pp/y for Cesar, Vindstat and Vattenfall, respectively.

Next, 20 randomly chosen units present in both Vindstat and Cesar were studied. For the units for which Cesar data are given for several WTs in aggregation, the corresponding data from Vindstat were also aggregated. Figure 14 shows metadata from the two different databases. The start dates agree well except for two units where the differences are over one year. A plausible explanation is that Cesar start dates for older units (built before the certificate system started) are not always correctly reported. The installed capacities are the same except for one unit. The coordinates differ somewhat, in one case with as much as 65 km. This is not surprising since the coordinates for some units were taken from the nearest village. This might make the LTC perform slightly worse, but will not systematically effect the trend estimates.

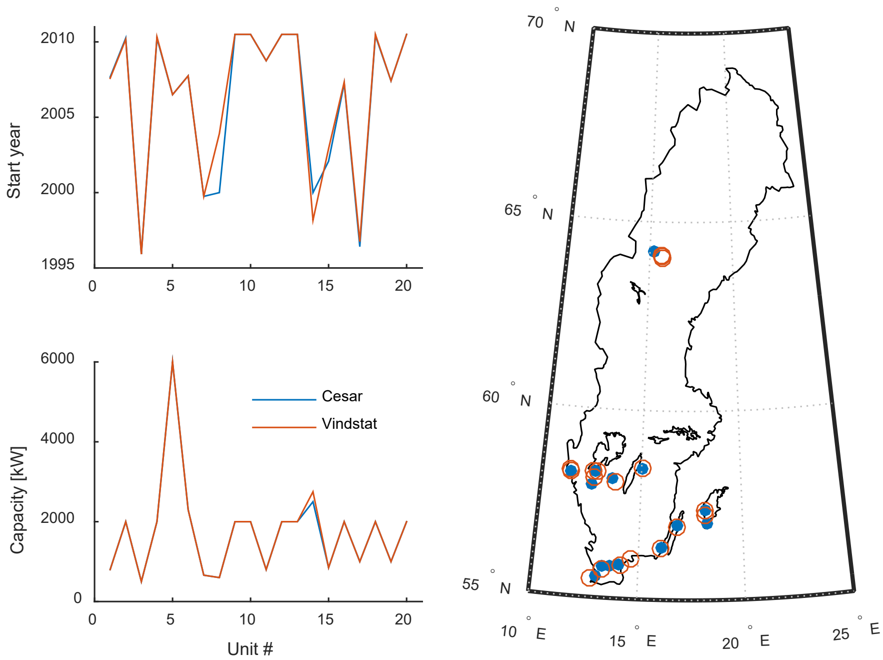


Figure 14. Comparison of metadata for 20 random units present in both databases.

Figure 15 shows monthly time series for all 20 units. In most cases, the CFs are identical or almost identical. For unit 1, 3 and 19, observations are missing from the Vindstat dataset. The curves for unit 14 are somewhat displaced since, as shown in Figure 14, the installed capacities differ in the two sets of metadata. Also note the constant production for unit 19 during around 20 months. As discussed in Section 3.2 in the main report, this must be taken into account in the LTC.

In conclusion, differences exist and the databases are not perfect. It is however hard to see that the main results would change much with perfect data since the detected errors and discrepancies can impact the trend estimates in both directions.

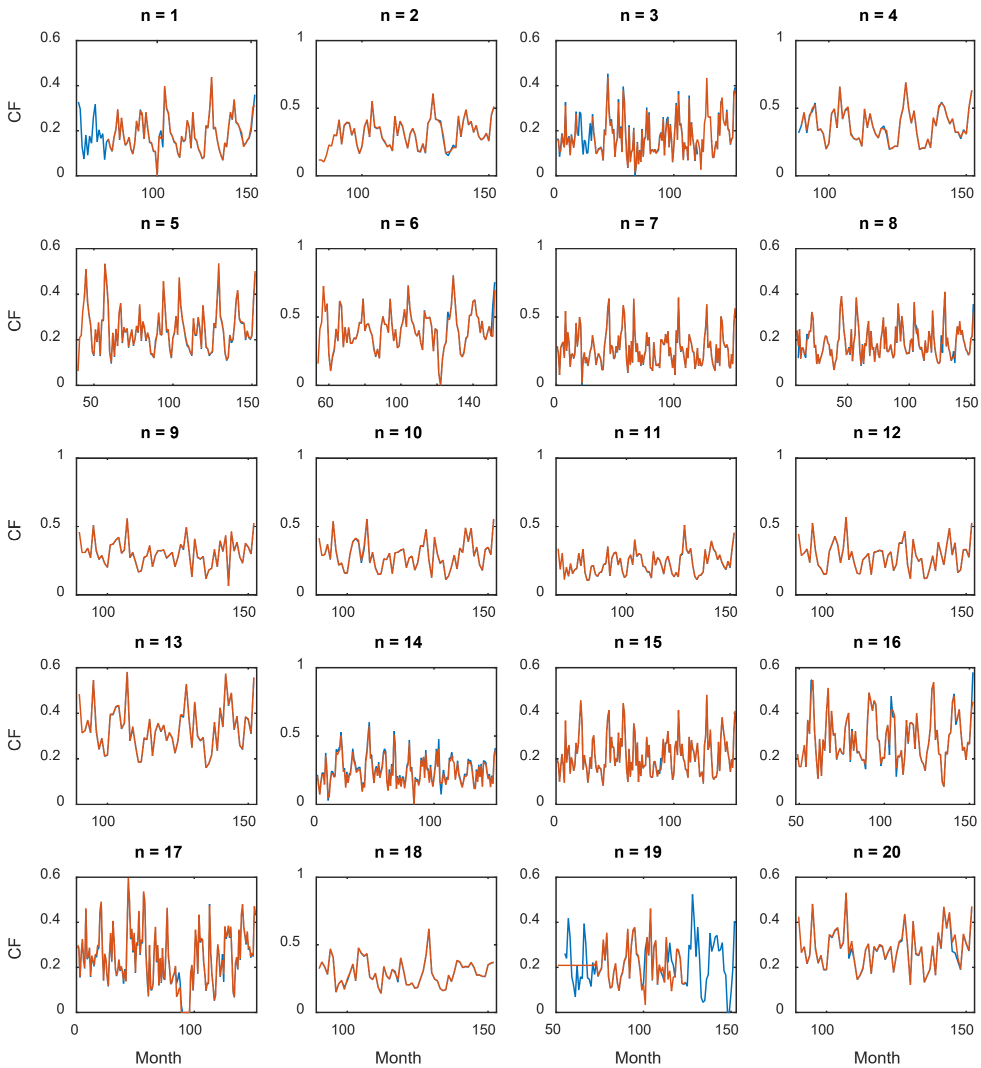


Figure 15. Comparison of raw, monthly capacity factors (CFs) in the Cesar (blue) and Vindstat (red) datasets for 20 random units. Month numbers on the horizontal axes are starting from May 2003.

4 Supplementary results

4.1 REANALYSES

Here, some additional analyses of the three different reanalyses are presented, see also Section 5.1 in the main report. First, standard deviations of seasonal patterns in CF_{LTC} (calendar month means) are compared to those for CF_{raw}. In average, the standard deviations for individual units were 28%, 36% and 24% for MERRA, ERA-I and ConWx relative those for the raw data. Most, but not all, of the seasonal patterns are thus removed with LTC. Seasonal patterns for all Vindstat units in aggregation are shown in Figure 16. The CF_{LTC}:s are highest for April, i.e. all reanalyses underestimates the generation for that month (recall that $CF_{LTC} = CF_{raw}/WI$).

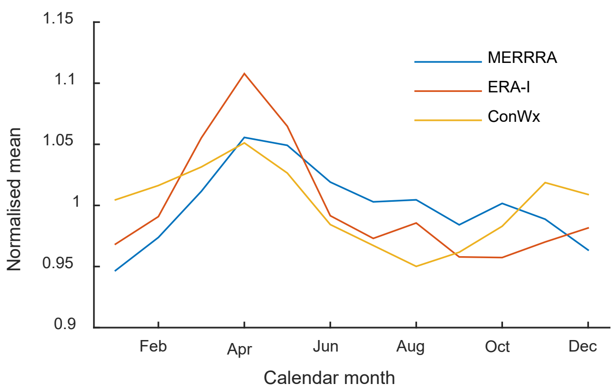


Figure 16. Normalised calendar month means of CF_{LTC} for all units in the Vindstat dataset. The normalization was done by dividing the calendar month means by the mean for the whole time series. Clearly, there exist (at least for some units) seasonal patterns also in the long-term corrected time series.

In a similar manner as for the calendar month means, the standard deviations of the high-frequency component of CF_{LTC} can be compared to those for CF_{raw}. A graphical example is given in Figure 17 and summary results in Table 2. The results in Table 2 are given both with and without the removal of very low CF_{LTC} observations.

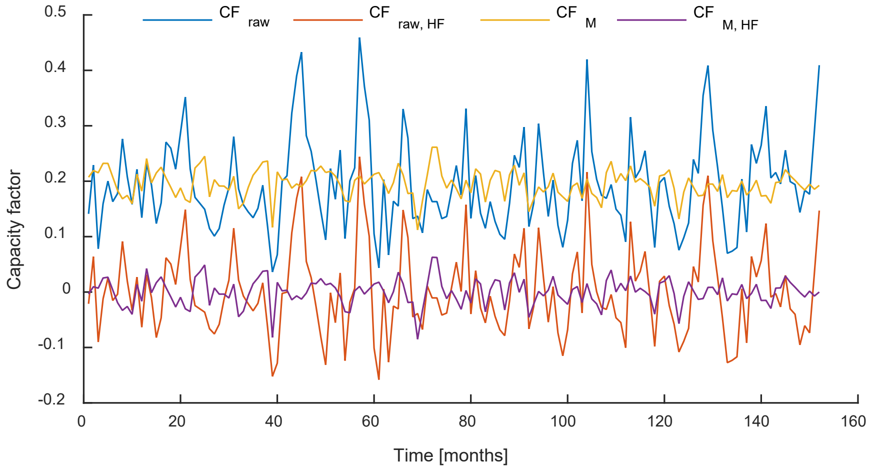


Figure 17. Example of raw and long-term corrected capacity factors (CF_{raw} and CF_{M}) and their corresponding high-frequency (subscript HF) components. In this particular case, the standard deviation of $CF_{M, HF}$ is 30% of that for $CF_{raw, HF}$.

Table 2. Average standard deviations of high-frequency part of CF_{LTC} relative ditto for CF_{raw}. Results are given both with and without removal of very low CF_{LTC} observations.

	MERRA	ERA-I	ConWx
Vindstat	43%	49%	41%
Cesar	44%	48%	39%
Vindstat (low observations removed)	32%	39%	29%
Cesar (low observations removed)	33%	39%	29%

A final note is that the power spectral density (PSD) estimates for ConWx fictive generation time series are clearly the most similar (in the high-frequency range) to the corresponding PSDs for measurements. This is however not very important for the current application.

4.2 REGRESSION DIAGNOSTICS

Two examples of regression diagnostics and a short discussion on the appropriateness of the linear model is given below. We begin by giving an example of a unit where the linear model with a seasonal component is more appropriate than the model without. As can be seen in the upper left panel in Figure 18, there is a clear pattern in CF_{LTC} and thus in the residuals when a seasonal component is not included. The residual autocorrelation function in the lower left panel has a sinusoidal pattern with high values around lag 12 and negative values for lags around 6 and 18. By using a seasonal component (right panels) the residuals are smaller in magnitude and more independent; no obvious pattern and consequently relatively low ACF for all lags. The ACF for lag 9 is close to the 2σ (around 95%) confidence bound, but this is exactly what can be expected for one out of twenty lags. The linear slopes (β_1) are -0.12 and -0.15 pp/y for the models with and without seasonal component respectively. There are strong reasons to believe that the latter is a better estimate.

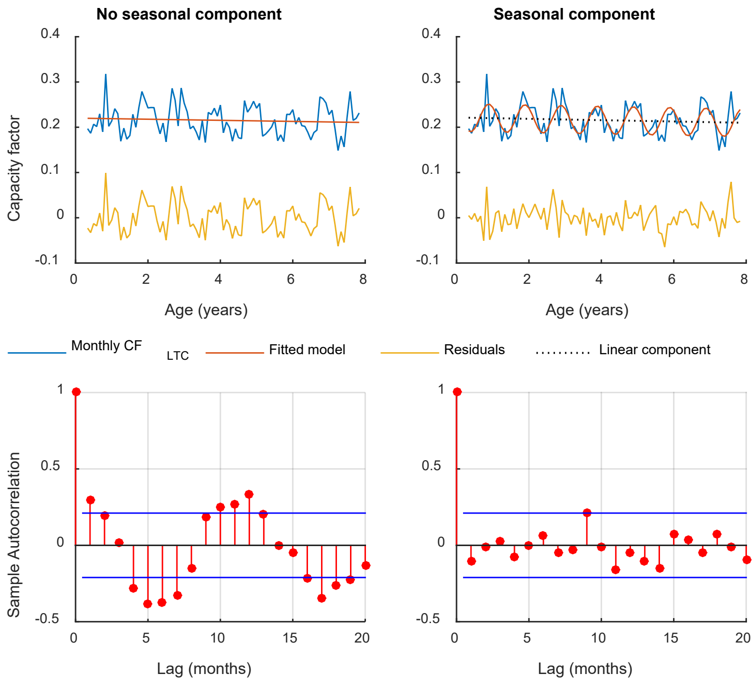


Figure 18. Selected regression diagnostics for a model without (left) and with (right) a seasonal component. Time series are given in the upper rows and sample ACF of residuals (with 2σ confidence bands) in the lower. In this case, a model with seasonal component is clearly more appropriate.

Figure 19 shows regression diagnostics for a unit with several consecutive months of zero generation. Such periods are, by far, the dominant explanation for poor diagnostics (when seasonal components are allowed, otherwise the residuals would be strongly autocorrelated for many units). The zero months have a strong influence on the fitted slope, i.e. strongly negative DFBETAs. Furthermore, the autocorrelation function has large values for low lags and the residuals are clearly not normally distributed.

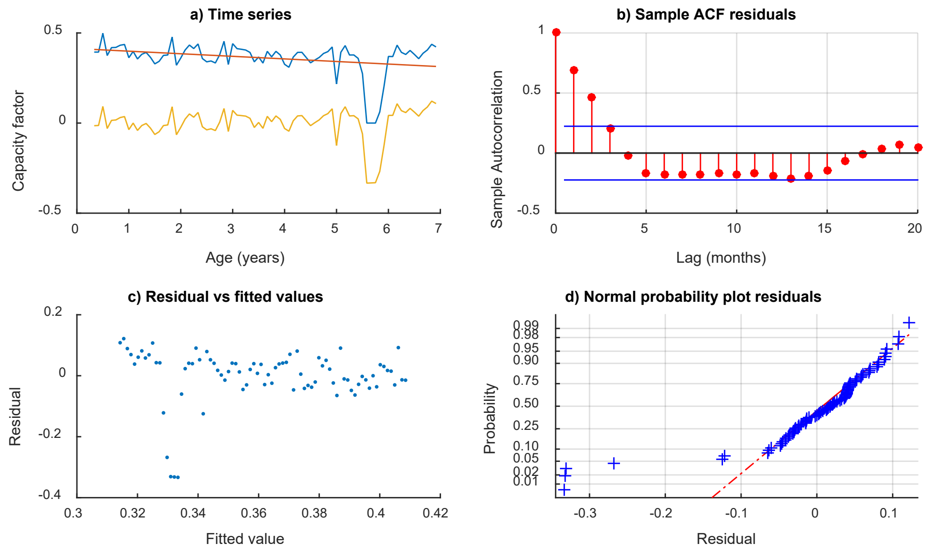


Figure 19. Typical example of error diagnostics for a unit where the linear model is not perfectly valid due to several consecutive zero observations. These observations correspond to strongly negative residuals and have a strong impact on the fitted slope parameter. (a) Time series of CF_{LTC} (blue), fitted model (red) and residuals (yellow); (b) Sample autocorrelation function of residuals including 2σ confidence intervals. The consecutive zero observations give rise to a significant ACF for low lags; (c) Residuals versus fitted values; (d) Normal probability plot of residuals.

A simple solution would be to consider the zero months as outliers and remove these observations. From a purely statistical point of view this might seem tempting, but would not be appropriate since the generation is actually zero and that increased downtime can be an important factor for performance decline which should be incorporated in the slope estimates. We thus accept dependence and non-linearity of the data for some of the units and calculate the linear trend anyway. As can be seen in Table 3, most units have 0-1 months with zero production. The violation of some of the assumptions for linear regression in some cases was however an important reason for also calculating slopes for cohorts and equivalent trends, see Section 4.2.2 in the main report. As will soon be shown, the results are very similar for all three methods and thus seem to be robust.

Table 3. Share of units with different number of months with zero production.

Number of zero months	0	1	2-5	6-10	10+
Vindstat	83%	10%	4%	1%	1%
Cesar	76%	13%	9%	2%	1%

High leverage is most often not an issue since we generally have equally spaced time series. In a few cases with long periods of missing data, the leverage can be relatively high before or after these periods. Such units were checked manually, but we did not remove any data points because of high leverage.

4.3 TABLES WITH MEAN AND MEDIAN TRENDS

In Table 4, mean and median trends including confidence intervals are given for linear regression for individual units (Method 1 as defined in Section 4.2.2 in the main report). Table 5 give the same information for equivalent trends. As described in the main document, a bootstrap method was used to calculate confidence intervals since the distributions of trends are generally heavy-tailed.

Tables 6-7 give results for normalised trends, see Section 4.2.3 in the main report.

Table 4. Mean and median trends (linear regression for individual units) with confidence intervals. The results are given in percentage points per year.

	Mean trend	Median trend
Vindstat, MERRA	-0.117 (-0.136, -0.099)	-0.102 (-0.115, -0.085)
Vindstat, ERA-I	-0.108 (-0.128, -0.088)	-0.127 (-0.146, -0.113)
Vindstat, ConWx	-0.092 (-0.112, -0.072)	-0.107 (-0.122, -0.092)
Cesar, MERRA	-0.111 (-0.139, -0.084)	-0.091 (-0.104, -0.074)
Cesar, ERA-I	-0.125 (-0.153, -0.097)	-0.122 (-0.137, -0.107)
Cesar, ConWx	-0.058 (-0.088, -0.029)	-0.057 (-0.071, -0.041)

Table 5. Mean and median equivalent trends with confidence intervals. The results are given in percentage points per year.

	Mean trend	Median trend
Vindstat, MERRA	-0.109 (-0.130, -0.086)	-0.096 (-0.108, -0.084)
Vindstat, ERA-I	-0.116 (-0.139, -0.093)	-0.134 (-0.147, -0.120)
Vindstat, ConWx	-0.098 (-0.120, -0.076)	-0.115 (-0.132, -0.102)
Cesar, MERRA	-0.089 (-0.118, -0.060)	-0.094 (-0.105, -0.072)
Cesar, ERA-I	-0.131 (-0.159, -0.102)	-0.162 (-0.181, -0.139)
Cesar, ConWx	-0.063 (-0.094, -0.034)	-0.105 (-0.124, -0.089)

Table 6. Mean and median normalised trends (linear regression for individual units) with confidence intervals. The results are given in percentage points per year. The normalization is done by dividing the trends by the average capacity factor for units with the same start year.

	Mean trend	Median trend
Vindstat, MERRA	-0.469 (-0.540, -0.396)	-0.391 (-0.452, -0.340)
Vindstat, ERA-I	-0.442 (-0.517, -0.368)	-0.513 (-0.584, -0.440)
Vindstat, ConWx	-0.396 (-0.470, -0.323)	-0.428 (-0.486, -0.367)
Cesar, MERRA	-0.455 (-0.561, -0.355)	-0.352 (-0.415, -0.297)
Cesar, ERA-I	-0.534 (-0.645, -0.431)	-0.493 (-0.557, -0.444)
Cesar, ConWx	-0.289 (-0.396, -0.185)	-0.236 (-0.289, -0.176)

Table 7. Mean and median equivalent, normalised trends with confidence intervals. The results are given in percentage points per year. The normalization is done by dividing the trends by the average capacity factor for units with the same start year.

	Mean trend	Median trend
Vindstat, MERRA	-0.423 (-0.506, -0.343)	-0.378 (-0.421, -0.325)
Vindstat, ERA-I	-0.455 (-0.538, -0.371)	-0.528 (-0.582, -0.467)
Vindstat, ConWx	-0.408 (-0.491, -0.325)	-0.476 (-0.531, -0.403)
Cesar, MERRA	-0.360 (-0.464, -0.254)	-0.370 (-0.437, -0.299)
Cesar, ERA-I	-0.548 (-0.657, -0.435)	-0.653 (-0.726, -0.556)
Cesar, ConWx	-0.314 (-0.424, -0.201)	-0.428 (-0.499, -0.349)

4.4 PERFORMANCE LOSS

In Section 5.6 in the main report, the evolution of the average performance factor for all Vindstat units was quantified. Figure 20 presents similar results for Vindstat units with different start years. Results are only given for ages with at least 60 monthly samples. Figure 21 shows a comparison of average performance factors for Vindstat and Cesar units. Only units with low-age observations (which are used for normalisation) are included. For Cesar, this implies that only 887 out of 1317 units were considered and that no high-age results are available.

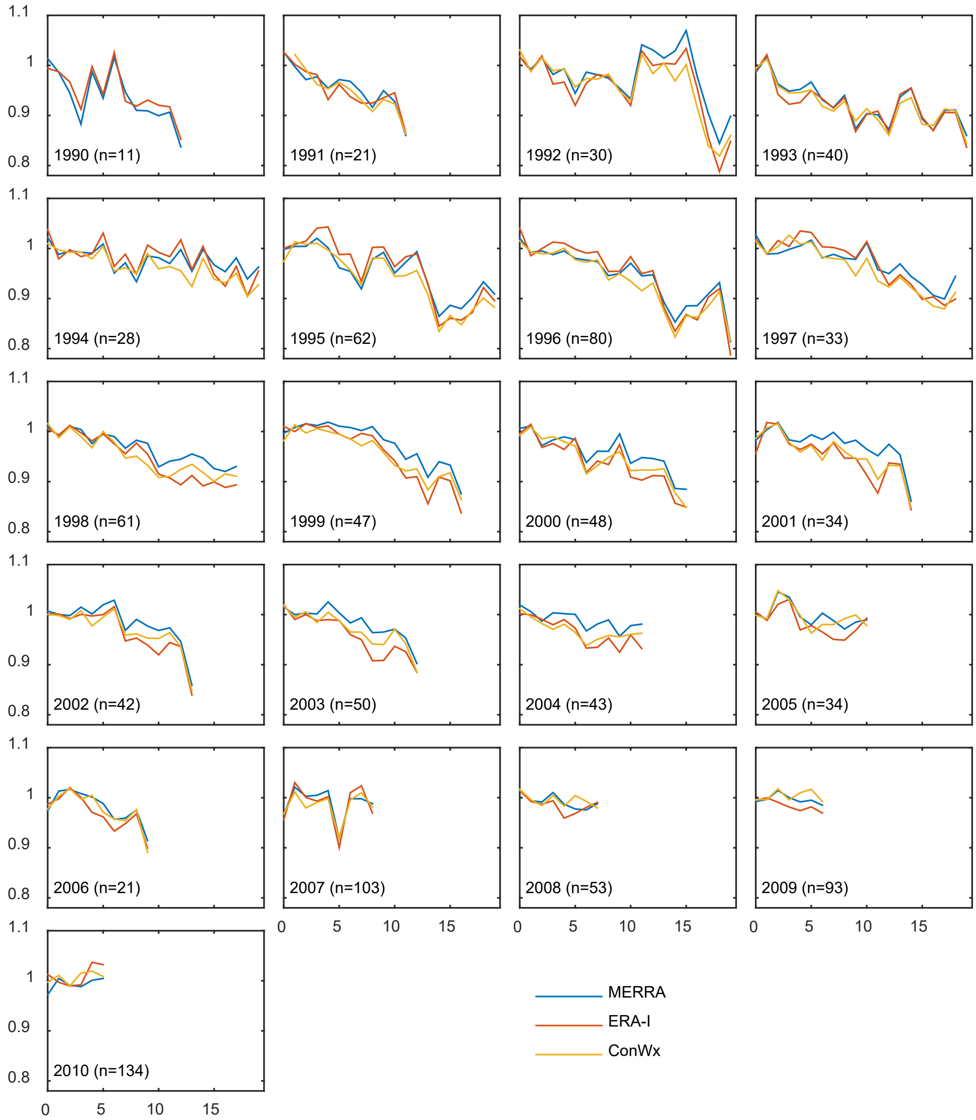


Figure 20. Average performance factors (vertical axes) depending on age (horizontal axel) for Vindstat units with certain start years. The start years and number of units are indicated in each panel. Note that for some ages, the number of samples is quite small.

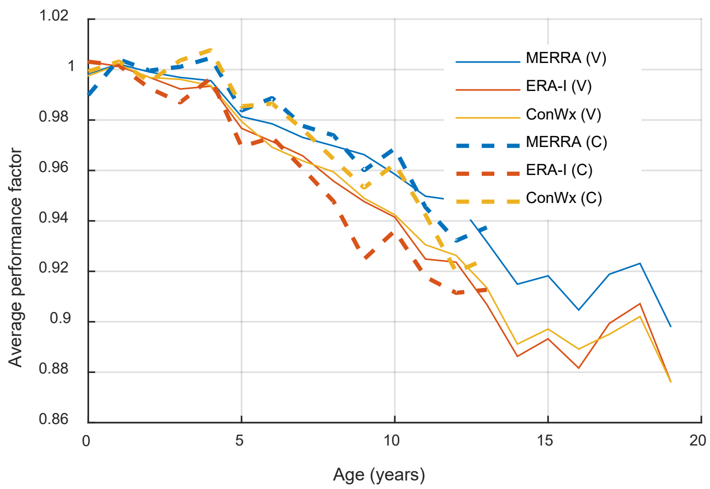


Figure 21. Comparison of average performance factors for Vindstat and Cesar units. Only units with low-age observations (which are used for normalisation) are included. V = Vindstat, C = Cesar.

4.5 INFLUENTIAL VARIABLES

In Section 5.7 in the main report, multiple linear regression models were employed in order to quantify the influence from several variables on the trends. Here, supplementary results are presented.

In Figure 22, coefficients for the models with selected linear terms are given. Figure 23 gives coefficients for separate models for each predictor variable, controlling for start year. The figures thus corresponds to methods 2 and 3 as defined in Section 4.2.5 in the main report.

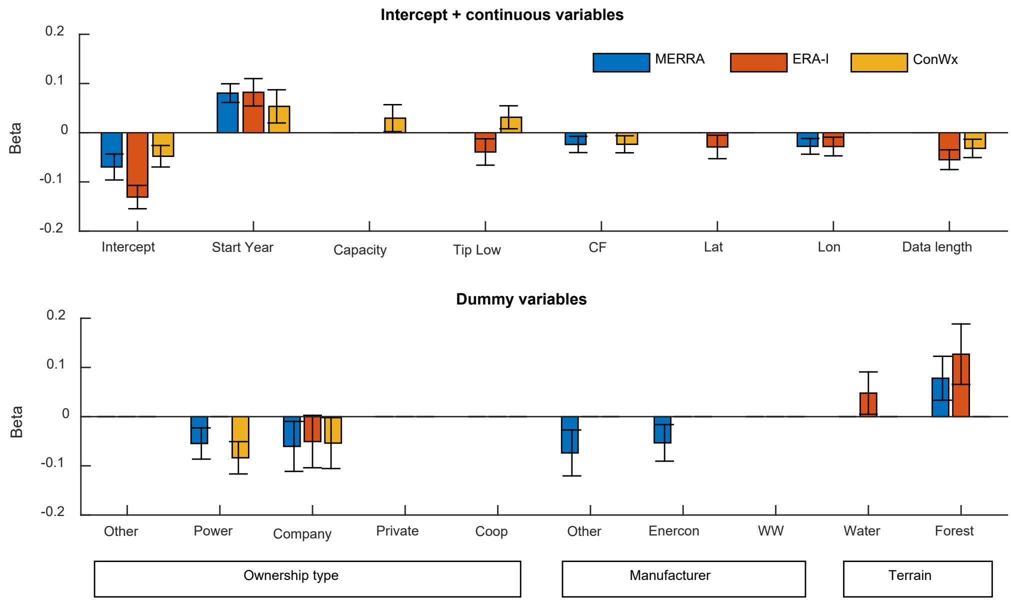


Figure 22. Coefficients in multiple linear regression model (selected linear terms but no interactions). Confidence intervals are also indicated.

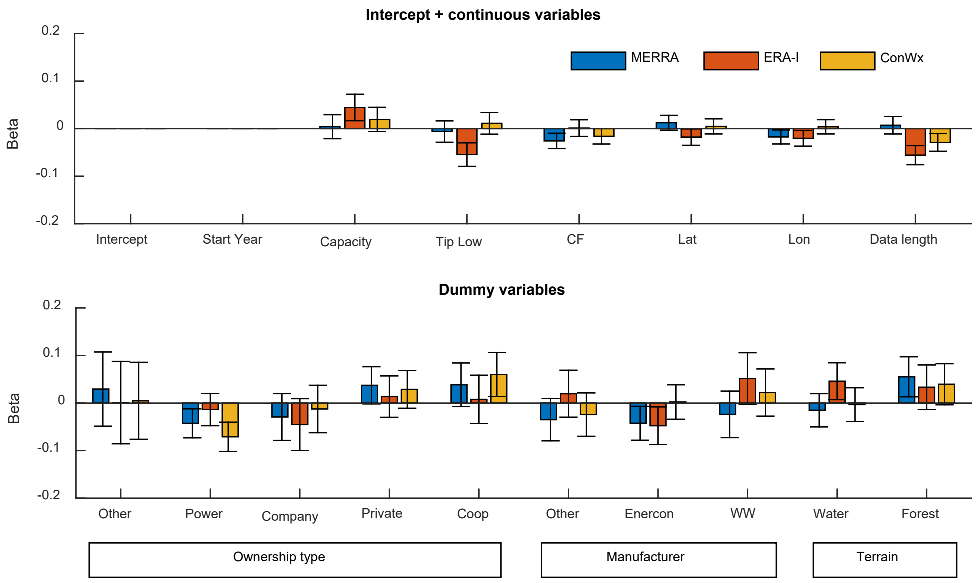


Figure 23. Coefficients in multiple linear regression model (separate models for each predictor variable, controlling for start year). Confidence intervals are also indicated.

Next, the full models with interaction terms are presented for MERRA (Figure 24), ERA-I (Figure 25) and ConWx (Figure 26). Note that the interaction terms have been orthogonalised, so the interpretation of the coefficients is not straight-forward. More details can be provided upon request.

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	-0.093425	0.010225	-9.1373	3.1537e-19
Start Year	0.049096	0.010201	4.8126	1.705e-06
CF	-0.016019	0.0090117	-1.7776	0.07576
Lon	-0.030696	0.0080298	-3.8227	0.00013965
Other manu	-0.15341	0.026346	-5.8229	7.6545e-09
Enercon	-0.046511	0.019405	-2.3968	0.016709
Forest	0.094452	0.023922	3.9484	8.3862e-05
Start Year:P	-0.031863	0.014615	-2.1802	0.029462
Start Year:Lat	-0.056765	0.020409	-2.7814	0.0055085
Start Year:Lon	0.02563	0.010877	2.3564	0.018631
Start Year:Enercon	0.16081	0.04956	3.2447	0.0012124
Start Year:Water	0.041634	0.020796	2.0021	0.045529
P:CF	0.032225	0.012505	2.5771	0.010098
P:Lat	0.035852	0.017597	2.0374	0.041862
P:Power	-0.055623	0.020494	-2.7141	0.0067521
P:Company	-0.11369	0.036565	-3.1092	0.0019257
Tip Low:CF	-0.051927	0.010926	-4.7527	2.283e-06
Tip Low:Data length	-0.056505	0.01265	-4.4669	8.7853e-06
Tip Low:Private	-0.058525	0.023603	-2.4795	0.013309
Tip Low:Coop	-0.084863	0.030047	-2.8243	0.0048269
Tip Low:Forest	-0.05186	0.025263	-2.0528	0.040333
CF:Lat	-0.015605	0.010006	-1.5595	0.11917
CF:Lon	0.023884	0.0091698	2.6046	0.0093251
CF:Power	0.032904	0.019157	1.7176	0.086162
CF:WW	-0.12817	0.027192	-4.7137	2.7551e-06
Lat:Lon	0.033041	0.007381	4.4764	8.4079e-06
Lat:Other manu	-0.059701	0.024999	-2.3882	0.017106
Lat:WW	-0.14094	0.076739	-1.8367	0.066539
Lon:Data length	0.028747	0.0087333	3.2916	0.0010289
Data length:Company	0.065425	0.025909	2.5252	0.011708
Data length:Other manu	0.095567	0.028241	3.384	0.00074042
Data length:Enercon	0.093073	0.044586	2.0875	0.03708
Other own:Water	-0.18399	0.089062	-2.0658	0.039086
Other own:Forest	0.25928	0.13006	1.9935	0.046462
Power:Water	-0.10857	0.041708	-2.6032	0.0093654
Company:Other manu	0.2905	0.17341	1.6752	0.094186
Company:Enercon	0.2021	0.05672	3.5631	0.00038269
Company:Forest	0.18449	0.092086	2.0035	0.045377
Private:Water	-0.16874	0.052844	-3.1932	0.0014481
Coop:Enercon	0.13472	0.061921	2.1756	0.029802

Number of observations: 1104, Error degrees of freedom: 1064
Root Mean Squared Error: 0.221
R-squared: 0.294, Adjusted R-Squared 0.268
F-statistic vs. constant model: 11.4, p-value = 8.92e-57

Figure 24. Multiple linear regression model with interaction terms (MERRA). The terms were chosen with a forward and backward selection procedure.

	Estimate	SE	tStat	pValue
(Intercept)	-0.053586	0.011599	-4.62	4.3111e-06
Start Year	0.049627	0.019126	2.5947	0.0095983
CF	-0.036706	0.00955	-3.8436	0.0001285
Lat	-0.039328	0.0097427	-4.0367	5.8132e-05
Data length	-0.10064	0.018821	-5.3475	1.0934e-07
Company	-0.058238	0.023857	-2.4412	0.014803
Other manu	-0.16947	0.030756	-5.5101	4.5074e-08
Enercon	-0.052259	0.022187	-2.3554	0.018684
WW	-0.059281	0.027932	-2.1223	0.034043
Start Year:Tip Low	-0.16342	0.03173	-5.1502	3.1039e-07
Start Year:Lat	-0.076377	0.014602	-5.2307	2.0363e-07
Start Year:Data length	-0.093812	0.016284	-5.7609	1.0977e-08
Start Year:Company	0.086514	0.050143	1.7253	0.084758
Start Year:Other manu	0.071058	0.036065	1.9703	0.049067
Start Year:Enercon	0.31446	0.079648	3.9482	8.398e-05
Start Year:WW	0.10604	0.06651	1.5944	0.11115
Start Year:Forest	-0.13334	0.034491	-3.8659	0.00011743
P:Tip Low	0.099946	0.020187	4.9511	8.59e-07
P:CF	0.039575	0.013574	2.9155	0.0036265
P:Lon	-0.048296	0.012733	-3.7931	0.00015727
P:Other own	-0.080925	0.038856	-2.0827	0.03752
P:Company	-0.11188	0.053721	-2.0827	0.037521
P:Enercon	-0.15078	0.031233	-4.8277	1.5857e-06
P:Water	0.065591	0.024529	2.6741	0.00761
Tip Low:CF	-0.084268	0.013491	-6.2463	6.0897e-10
Tip Low:Lon	0.077156	0.013387	5.7634	1.0821e-08
Tip Low:Data length	-0.07468	0.025831	-2.8911	0.0039178
CF:Power	0.048865	0.018564	2.6322	0.0086081
CF:Private	0.040262	0.023515	1.7122	0.087154
CF:Enercon	0.125	0.024469	5.1086	3.8516e-07
CF:WW	-0.12014	0.028968	-4.1472	3.637e-05
CF:Water	0.028675	0.019648	1.4594	0.14474
Lat:Lon	0.055431	0.010797	5.1342	3.3743e-07
Lat:Power	0.029743	0.020061	1.4826	0.13848
Lat:Coop	-0.088385	0.026861	-3.2904	0.0010335
Lat:Other manu	-0.10205	0.025871	-3.9446	8.5218e-05
Lat:WW	-0.27113	0.087054	-3.1145	0.0018921
Lat:Water	-0.14459	0.029653	-4.876	1.2493e-06
Lat:Forest	0.061448	0.029286	2.0982	0.036124
Lon:Power	0.032607	0.017711	1.8411	0.065888
Lon:Enercon	-0.083953	0.02444	-3.435	0.00061569
Lon:Forest	-0.17084	0.0313	-5.4583	5.9939e-08
Data length:Company	0.084075	0.028144	2.9874	0.0028792
Data length:Other manu	0.21001	0.035669	5.8877	5.2593e-09
Data length:Enercon	0.12941	0.060791	2.1287	0.033509
Data length:Water	0.058075	0.019691	2.9493	0.0032554
Company:Other manu	0.5642	0.18041	3.1273	0.0018125
Company:Enercon	0.11427	0.067328	1.6973	0.089937
Company:Water	0.23388	0.063662	3.6738	0.00025101
Coop:Water	0.23325	0.054226	4.3015	1.8543e-05
Enercon:Water	-0.14177	0.059693	-2.3749	0.017732

Number of observations: 1104, Error degrees of freedom: 1053
Root Mean Squared Error: 0.228
R-squared: 0.414, Adjusted R-Squared 0.386
F-statistic vs. constant model: 14.9, p-value = 3.53e-90

Figure 25. Multiple linear regression model with interaction terms (ERA-I). The terms were chosen with a forward and backward selection procedure.

	Estimate	SE	tStat	pValue
(Intercept)	-0.046968	0.009837	-4.7746	2.055e-06
Start Year	0.052866	0.020341	2.599	0.0094797
P	0.027974	0.018121	1.5437	0.12295
CF	-0.02597	0.0092761	-2.7996	0.0052098
Lat	-0.029718	0.0090261	-3.2925	0.0010261
Data length	-0.053159	0.015579	-3.4123	0.00066878
Power	-0.037757	0.015156	-2.4913	0.012881
Other manu	-0.20937	0.031187	-6.7136	3.1001e-11
Start Year:Tip Low	-0.070094	0.028799	-2.4339	0.015104
Start Year:Lat	-0.09297	0.019306	-4.8155	1.6834e-06
Start Year:Data length	-0.037155	0.015501	-2.3969	0.016709
Start Year:Enercon	0.18227	0.03761	4.8463	1.4471e-06
Start Year:Forest	-0.14628	0.047926	-3.0521	0.0023292
P:Tip Low	0.083014	0.020157	4.1184	4.1142e-05
P:CF	0.034379	0.012044	2.8544	0.0043966
P:Lon	-0.04063	0.012392	-3.2788	0.0010766
P:Other own	-0.14018	0.048703	-2.8783	0.0040795
P:Company	-0.097246	0.035841	-2.7133	0.0067708
P:Other manu	-0.091219	0.03388	-2.6924	0.0072067
P:Enercon	-0.15602	0.033207	-4.6985	2.9686e-06
P:Water	0.094912	0.022783	4.1659	3.3551e-05
P:Forest	0.092399	0.041432	2.2302	0.025947
Tip Low:CF	-0.069347	0.012493	-5.5507	3.6005e-08
Tip Low:Lon	0.061965	0.011289	5.4889	5.0693e-08
Tip Low:Data length	-0.03692	0.021995	-1.6786	0.093524
Tip Low:Other own	0.080581	0.042587	1.8921	0.058747
CF:Enercon	0.089017	0.023236	3.831	0.00013515
CF:WW	-0.1044	0.026445	-3.9477	8.4145e-05
Lat:Lon	0.069866	0.010798	6.4705	1.4942e-10
Lat:Data length	-0.048272	0.016623	-2.9039	0.0037627
Lat:Power	-0.081243	0.018985	-4.2793	2.0462e-05
Lat:Coop	-0.04723	0.024755	-1.9079	0.056674
Lat:WW	-0.24044	0.074286	-3.2366	0.0012473
Lat:Water	-0.19146	0.030392	-6.2996	4.3784e-10
Lat:Forest	-0.099131	0.030028	-3.3013	0.00099471
Data length:Company	0.055092	0.025881	2.1287	0.033513
Data length:Other manu	0.10052	0.033528	2.9981	0.0027805
Data length:Water	0.050076	0.018471	2.7111	0.0068153
Company:Other manu	1.2608	0.1797	7.0161	4.0804e-12
Company:Enercon	0.2022	0.056301	3.5915	0.00034398
Company:Water	0.1415	0.057959	2.4414	0.014795
Company:Forest	-0.17639	0.094354	-1.8694	0.061844
Private:Forest	-0.1448	0.064304	-2.2518	0.02454
Coop:Water	0.21623	0.051094	4.2321	2.5172e-05
Other manu:Forest	-0.27204	0.081659	-3.3314	0.00089419

Number of observations: 1099, Error degrees of freedom: 1052
Root Mean Squared Error: 0.215
R-squared: 0.408, Adjusted R-Squared 0.382
F-statistic vs. constant model: 15.7, p-value = 6.79e-90

Figure 26. Multiple linear regression model with interaction terms (ConWx). The terms were chosen with a forward and backward selection procedure.

4.6 DOWNTIME

Figure 27 gives more detailed information on the seasonal downtime patterns for Cesar units of different latitudes. Table 8 shows how the mean and median trends for all Cesar units change if downtime data is excluded.

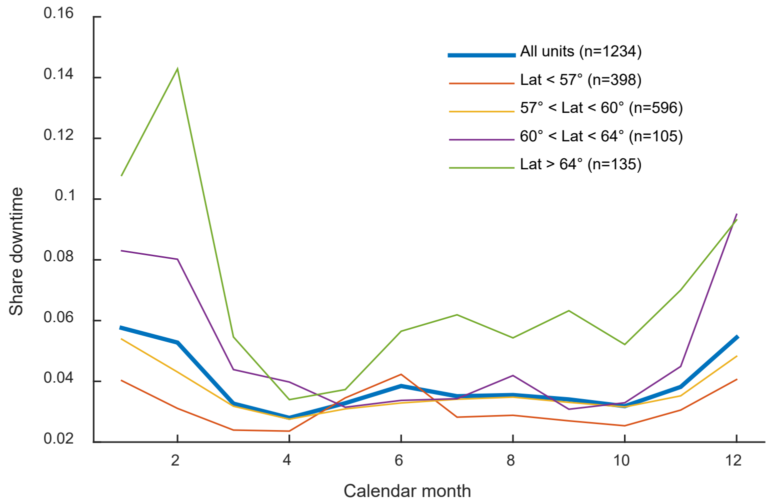


Figure 27. The downtime is generally larger during winter, especially for units in the north. For all data (age 0.33 – 20 years), the estimated downtime is 4.0%.

Table 8. Impact on mean and median trends by excluding downtime data. All results are given in percentage points per year for Cesar data with units with more than one wind turbine excluded.

	Mean (downtime included)	Mean (downtime excluded)	Median (downtime included)	Median (downtime excluded)
MERRA	-0.11	-0.05	-0.09	-0.05
ERA-I	-0.12	-0.06	-0.12	-0.09
ConWx	-0.05	+0.01	-0.05	-0.04
Average	**-0.09**	**-0.03**	**-0.09**	**-0.06**

Figure 28, shows results for cohort regressions (Method 2 as described in Section 4.2.2 in the main report) with and without downtime data. Especially for older units, a relatively large part of the performance decline can be attributed to increased downtime.

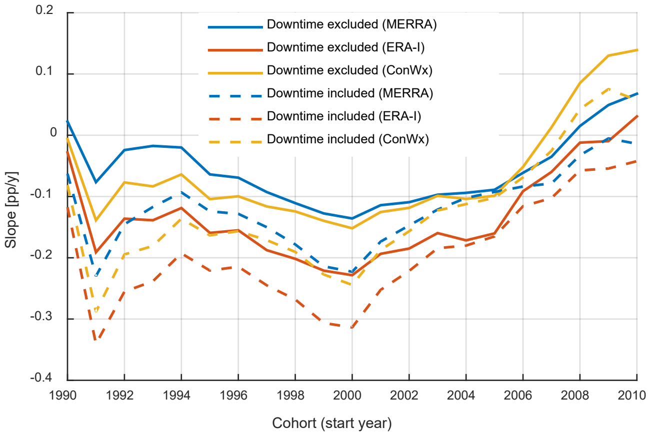


Figure 28. Impact on trends of different cohorts from excluding or including downtime (Cesar data, only units corresponding to one wind turbine).

WIND TURBINE PERFORMANCE DECLINE IN SWEDEN

The main objective here is to answer how much wind turbine performance declines with age. This question is of great importance for the profitability of wind farms and affects the required installed capacity to fulfil renewable energy targets.

Previous large-scale studies of this type are however very scarce. During the first years of operation, the production is nearly constant, but subsequently it begins to decline. Wind turbines constructed before 2007 lose around 0.15 capacity factor percentage points per year in absolute terms, corresponding to a life-time energy loss of 6 %.

A gradual increase of downtime accounts for around 1/3 of the decline and worsened efficiency for the rest. In comparison to results from the UK, Swedish wind farms deteriorate much slower.

Energiforsk is the Swedish Energy Research Centre – an industrially owned body dedicated to meeting the common energy challenges faced by industries, authorities and society. Our vision is to be hub of Swedish energy research and our mission is to make the world of energy smarter!

Energiforsk