Millar, Robert John; Ekström, Jussi; Lehtonen, Matti; Koivisto, Matti; Saarijärvi, Eero; Degefa, Merkebu

Probabilistic prosumer node modeling for estimating planning parameters in distribution networks with renewable energy sources

Published in:
Proceedings of the 58th IEEE International Scientific Conference on Power and Electrical Engineering of Riga Technical University, RTUCON 2017

DOI:
10.1109/RTUCON.2017.8124833

Published: 01/01/2017

Document Version
Peer reviewed version

Please cite the original version:
Millar, R. J., Ekström, J., Lehtonen, M., Koivisto, M., Saarijärvi, E., & Degefa, M. (2017). Probabilistic prosumer node modeling for estimating planning parameters in distribution networks with renewable energy sources. In *Proceedings of the 58th IEEE International Scientific Conference on Power and Electrical Engineering of Riga Technical University, RTUCON 2017* IEEE. https://doi.org/10.1109/RTUCON.2017.8124833

Probabilistic Prosumer Node Modeling for Estimating Planning Parameters in Distribution Networks with Renewable Energy Sources

Robert John Millar
Jussi Ekström
Matti Lehtonen
Aalto University
Espoo, Finland
john.millar@aalto.fi

Matti Koivisto
Danish University of Technology
Denmark

Eero Saarijärvi
Trimble Oy
Espoo, Finland

Merkebu Degefa
SINTEF Energi AF
Trondheim, Norway

***Abstract*— With the increase in distributed generation, the demand-only nature of many secondary substation nodes in medium voltage networks is becoming a mix of temporally varying consumption and generation with significant stochastic components. Traditional planning, however, has often assumed that the maximum demands of all connected substations are fully coincident, and in cases where there is local generation, the conditions of maximum consumption and minimum generation, and maximum generation and minimum consumption are checked, again assuming unity coincidence.**

Statistical modelling is used in this paper to produce network solutions that optimize investment, running and interruption costs, assessed from a societal perspective. The decoupled utilization of expected consumption profiles and stochastic generation models enables a more detailed estimation of the driving parameters using the Monte Carlo simulation method.

A planning algorithm that optimally places backup connections and three layers of switching has, for real-scale distribution networks, to make millions of iterations within iterations to form a solution, and therefore cannot computationally afford millions of parallel load flows in each iteration. The interface that decouples the full statistical modelling of the combinatorial challenge of prosumer nodes with such a planning algorithm is the main offering of this paper.

Keywords—Distributed Generation, Distribution Network Planning, Monte Carlo Simulation, Statistical Load Analysis, Wind generation Analysis

I. INTRODUCTION

Whilst still a fertile topic for research, electricity consumption modelling is a mature discipline in power systems, and it is reasonable to assert that when there are tens or more customers behind a given substation, load profiles can be estimated with a reasonable degree of certainty [1]-[5]. AMR measurements give the potential to further refine load profiling for planning purposes, although the demand of single and small numbers of low voltage (LV) customers will always involve a high degree of uncertainty [6]. On the other hand, the generation time series from renewable sources such as wind and solar are highly stochastic, and are challenging to predict.

Depending on viewpoint, the traditional assumption that the extreme conditions (either maximum demand / minimum generation or maximum generation / minimum demand) are fully coincident at the secondary substation level may be considered pragmatic, but is more likely to be considered overly conservative. The judicious use of coincidence factors reduces the margin, but with the rapid increase in distributed generation connected at the medium (MV) or low voltage (LV) level, the analysis becomes far more complex, requiring the time dependent modelling of each generation unit and consumption type. Once the consumption and generation has been modelled, the uncertainty in the stochastic models can be modelled using the Monte Carlo simulation method [6]-[8].

This paper shows that the Monte Carlo method, with a selected confidence interval (CI), can be used to produce safe parameterization of the network. What is considered safe, however, is ultimately the network company's choice, which may be expressed as a CI, e.g., the 95th percentile. We have used the Monte Carlo simulation method to analyse wind generation, including both spatial and temporal dependencies (the modelling was carried out as shown in [7], [8]). Its high stochasticity means that using only the maximum generation (i.e., the installed capacity) to model wind generation is clearly not representative of the observed behaviour (see the red simulated time series in Fig. 1).

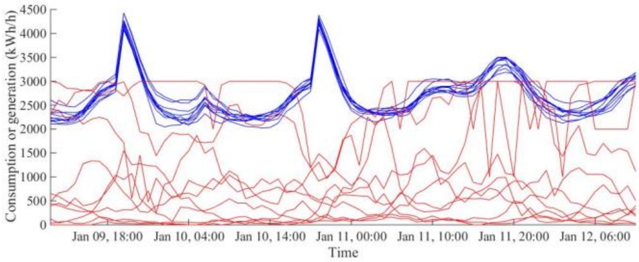

Figure 1 Ten Monte Carlo simulation runs of aggregate consumption (blue) and wind generation (red) showing the strong variability in wind generation and the relative lack of variability (around the expected path) in aggregated consumption (the wind and consumption scenarios are as specified in [6]).

When, however, enough customers are clustered together, as they often are in zoned urban, suburban and rural environments, the aggregate consumption is relatively predictable [5], [6]. In Monte Carlo simulation, this means that the individual simulation runs are close to the expected path (the blue profiles in Fig 1). The expected path (i.e., the profile) is dictated by outside temperature and other exogenous variables [6]. Therefore, in the context of a network planning paper that utilises a large amount of parameterisation with questionable accuracy, the consumption in this paper was modelled using only consumption profiles (i.e., the expected paths). However, wind generation is modelled using a stochastic model, with 100 Monte Carlo runs (with hourly resolution).

The reliability analysis has not yet been extended to cover the islanded supply of other secondary substation loads during grid faults. The ability of a substation area to supply itself is considered, but not the ability to supply a wider area in times when local production is high enough to do so. Switch placement in distribution networks with distributed generation has been treated, for example, by [9]. It should also be noted that this paper does not utilise the geographical interface developed in [10], [11], in order to better highlight the topological effect of the time series analyses we conduct in this paper.

Load growth can at present be node specific, but the decoupling of local generation from demand growth will also have to be implemented via a third dimension in the nodal aggregate time series simulation data, which has not yet been implemented. To retain a reasonable level of confidence when consumption and operational considerations are also fully modelled will require much larger data sets, containing thousands of Monte Carlo simulation runs (times the 40 or so years needed to represent load and generation growth up to the planning horizon).

Running a multitude of parallel power flows at every iteration in a distribution planning algorithm is computationally too heavy. Therefore, this paper proposes a decoupled approach, whereby node specific *coincidence factors* and the utilization times for losses (*loss times*) are computed and then periodically updated by pausing the network planning iteration process at critical stages in the algorithm. For a given network topology, power flows are run with hourly resolution over a multitude of Monte Carlo simulated realizations of a scenario year, such that worst (for technical constraints with a given confidence interval) and load-squared average (for losses) parameterization can be reliably achieved.

The two main interface parameters couple the statistical analysis with the planning algorithm are the coincidence factors and loss times. A coincidence factor for a particular node in the network is defined as the maximum (95 percentile confidence) hourly power flow in its feeding line section from the 876000 simulated hours of Monte Carlo data, divided by the summation of the line power flow that would occur if all the maximum nodal demands would occur simultaneously. The loss time is defined as the equivalent number of hours of maximum power flow in a line section that would give the same losses that typically occur in a year (8760 h).

The coincident factors, f_{coinc}, and loss times, T_{losses}, are used in the single shot backward sweep power flows based on maximum and minimum demand (where the minimum at some nodes may be negative, implying net generation) and minimum acceptable network voltage during intermediate iterations of the planning algorithm. Naturally, these two sets of parameters change as the topology of the network changes during the planning process, hence they must be re-computed as the network evolves towards the optimum solution.

We have earlier implemented Microgrid parameterization of secondary substation areas, [12], [13]. Microgrids, however, constitute a very specific and well-behaved type of prosumer node. By definition, they can *take care of themselves*, i.e. seamlessly disconnect and reconnect to the main MV or LV grid, and take care of their own demand, with a combination of DG, storage and load shedding. The prosumer node modelling used in this paper makes no such assumptions, and is therefore a useful improvement on previous work.

With this decoupled approach, the network is coerced towards the theoretical optimum by network modifying functions [14]-[16], noting that our approach is only one of many, e.g. [17], [18]. The statistical modelling and the decoupled interface are the major new contributions in this paper, but we must confess that coupling a full probabilistic modelling of consumption and generation with a network planning algorithm involves some compromises.

Section III illustrates the methodology covered in Section II with MV network plans that combine MV connected wind turbines alongside various consumption types aggregated to the secondary (MV/LV) substation level, with and without the probabilistic modelling.

II. Methodology

The overriding network optimisation is to produce a societal cost minimum network, where the costs include investments, operation and maintenance, loss, and interruption costs, subject to the thermal and voltage related technical restraints, and is described, for example, in [12].

This section commences with a description of the wind generation modelling necessary in forming the net-demand prosumer time series data.

A. *Wind Power Monte Carlo simulation*

The method used in this paper to generate the wind power time series consists of the transformed ARC model (Autoregressive model utilizing a spatial Correlation matrix) to generate the wind speed time series and a turbine model to convert the wind speed time series to the power domain [7], [8]. With this approach, 100 hourly one year time series (possible realizations of the wind speeds) are simulated for 8 wind turbines in specific locations.

The procedure utilized to generate the multivariate temporally and spatially correlated wind speed time series is as follows. First, univariate normally distributed time series are simulated for each turbine location using an autoregressive (AR) model, which can be written for location *i* as

$$z_t = c + \sum_{i=1}^{p} a_i z_{t-i} + e\,, \qquad (1)$$

where $a_1,\ldots, a_p$ are the AR model parameters, *p* is the order of the AR model, *c* is the intercept term and *e* is the error term of the model. Second, the monthly changing diurnal structures, estimated from two high and 19 low altitude locations in Finland are added to the time series, as presented in [8]. Third, the spatial correlations are added to the simulated data with the Cholesky decomposition by linking the distances between the turbine locations to the correlations between those locations, as presented in [7], [8]. Finally, the simulated normally distributed time series are transformed to wind speed time series through probability integral transformations. The marginal distributions utilized to obtain the wind speed time series through the transformations are considered to be Weibull distributed, as in [8]. The Weibull parameters defining the marginal distributions are obtained from the Finnish Wind Atlas database [19] according to the coordinates of each location.

In the last phase of the wind power simulation procedure, the turbine model presented in [8] is used to transform the wind speed time series to the power time series according to the specification of the turbines. For four out of the 8 locations, Winwind WWD-1-56 1 MW turbines [20] are considered and for the rest of the locations Enercon E-12 30 kW turbines [21] are utilized.

B. *Aggregation to Prosumer Nodes and Parameterisation for Network Planning Algorithm*

Fig. 2 gives an overview of the planning algorithm. If time series data is available, the pre-processing of prosumer (which may be demand-only, generation-only, or both) data is needed to allow the MV topology planning algorithm to produce a good initial network. This involves establishing, for each node *n* (secondary substation or MV connection point) in the network, the net demand time series *P*[*n*, *t*], *Q*[*n*, *t*], where, in this paper, *t* = (1h,...,8760 h). Fig. 3 shows the net active power demand *P*[*n*, *t*] for a single (8760 h) simulation run in the top figure. The lower figure illustrates the risk of using only a single simulation run's time series (whether based on measurements or modelled) as, for the particular simulated year in the top figure, the maximum would be well below the maximum for the 100 simulations of a modelled year or even the 95th CI. Therefore, we have 100 simulation runs for both *P*[*n*, *t*] and *Q*[*n*, *t*], for all *n*. The 95th CIs for the maximum and minimum *P*[*n*, *t*] and *Q*[*n*, *t*] for each node are calculated from the maximums and minimums of the 100 simulated runs. e.g., for the maximum active demand for node *n*, this means the 95th CI of max(*P*(*n*,*t*)), as detailed in [6].

When net demand goes negative, the node is acting as a net generator (this can be seen for some hours in Fig. 3). The net active demand for node *n* is thus defined as

$$P(n,t) = \sum_{d=1}^{D} P_{\text{dem}}(d,t) + \sum_{g=1}^{G} P_{\text{gen}}(g,t)\,, \qquad (2)$$

where *d* represents the demand type (district heating, direct electric heating, storage heating, etc.) and *g* the generators behind the node in question, noting that generation is mathematically treated as negative demand.

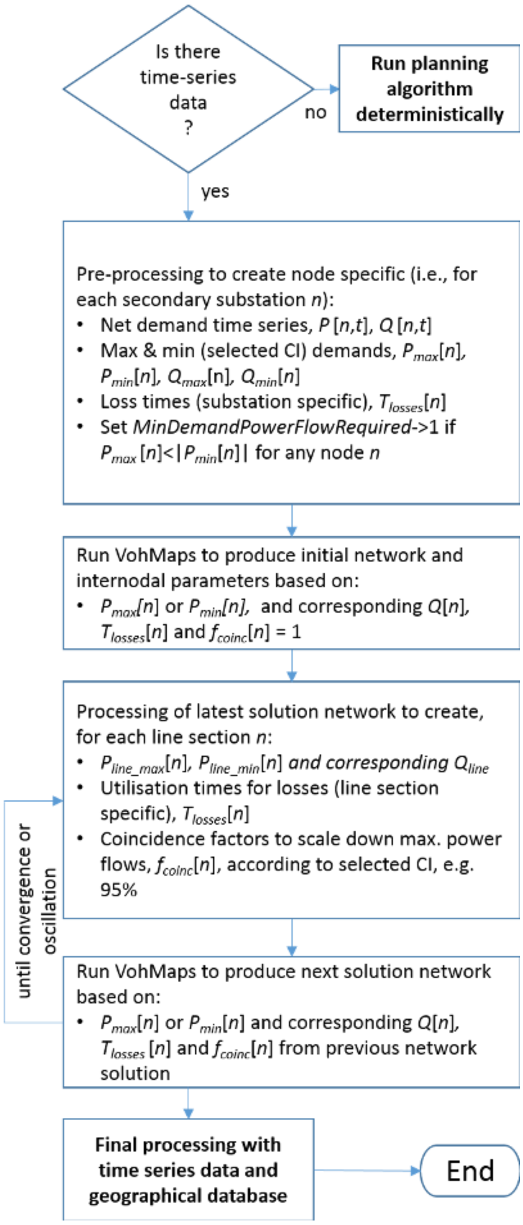

Figure 2 Overview of additional stages required when time series data is utilised in the MV topology planning algorithm. Note that *n* represents both the node *n* and the line section that feeds node *n*.

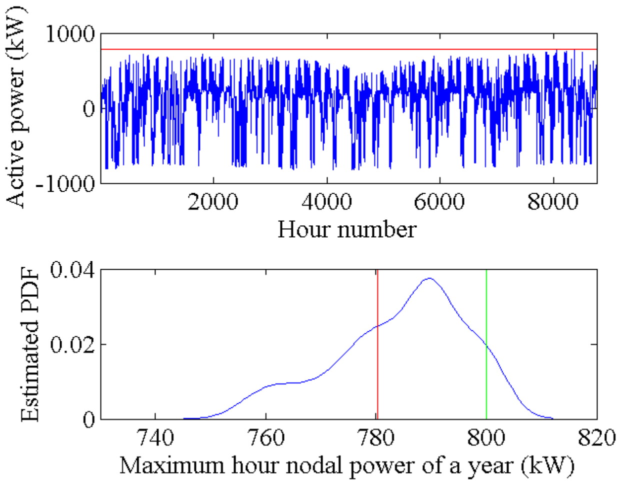

Figure 3 Active power demand of a prosumer node (node 3 in Table II in the Appendix). The top figure shows one simulation run with its maximum shown in red. The bottom figure shows the estimated PDF of all (100) simulation runs, with the maximum from the single run in the top figure shown in red and the 95th CI shown in green.

This is calculated for all the Monte Carlo simulation runs. The resulting 100 max($P(n,t)$) values are used to calculate the required CIs for the maximum net active power demand and, similarly, for the minimum net active and corresponding reactive power demands. The maximum net active power demand is abbreviated $P_{max}(n)$, which corresponds to the selected CI of the max($P(n,t)$). Similar symbols are used for the minimum net active power and the corresponding reactive power demands.

The net demand time series for each node or upstream (feeding) line section n, the loss times, $T_{\text{losses}}(n)$, are calculated from:

$$T_{\text{losses}}(n) = \frac{\sum_{t=1}^{8760} \left(P(n,t) + Q(n,t)\right)^2}{\left(\max\left(S(n,t)\right)\right)^2} \tag{3}$$

Equation (3) defines T_{losses}(n) for one simulation run. The T_{losses}(n) used in the planning algorithm is the average of the 100 T_{losses}(n) values. The coincidence factors for the initial network are set to unity, which is conservative.

C. The Context: A Distribution Network Planning Algorithm

The initial network is the starting point for the first round of time series modelling for each line section and the network modifying functions, but also allows initial parameterising of the geographical installation and operational environment for the network, which is not covered in this paper. A typical network solution is given for this and the subsequent stages of the algorithm in Section III.

The net demand time series are the key parameters for the planning, along with substation specific interruption costs (load averaged €/kW/permanent fault, €/kW/short interruption, and €/kWh of energy not supplied), node specific net growth, expressed as per unit values based on the first year for every year up to the planning time horizon, and the basic cost parameters, such as investment costs, loss costs and the electrotechnical parameters of the network components, see the Appendix. Naturally, all existing network must be parameterized, but the network plans shown in this paper are Greenfield, to better show the impact of the two main planning philosophies (deterministic vs. probabilistic), whereas in the real world, the algorithm is mostly used for Brownfield planning.

The interface between a full statistical treatment of nodal consumption and distributed generation is the main challenge addressed in this paper. The traditional methodology of assuming that the maximum nodal demands occur simultaneously in the network is now modified via periodically computed coincidence factors based on a Monte Carlo simulation analysis of the nodal time series and how they sum to each line section n in the radially operated network.

$$f_{\text{coinc,single\_run}}(n) = \frac{\max\left(\sum_{t=1}^{8760} S_{\text{line}}(n,t)\right)}{\sum_{i=1}^{I}\left(\max\left(S(i,t)\right)\right)} \tag{4}$$

where i are the downstream nodes fed by line section n. Equation (4) yields as many coincidence factors for every line section as there are Monte Carlo simulation runs. These 100 values per line section n are analysed in a similar manner as max($P(n,t)$) above, to give the required CI for $f_{coinc}[n]$. The coincidence factors are then used in the next iterations of the planning algorithm to scale down the line power flows based on full coincidence, which themselves are 95 percentiles of the simulation run maximums. Line specific loss times are calculated using line power flows in an equation similar to (3).

Running the relevant parts of the MV topology planning algorithm can be controlled by *settings* variables. A *sprint* version utilises a sector-based initial network generation [15] followed by a user-stipulated time-limited number of basic branch exchanges and polishing functions. Because the sample network section shown in the Results section is rather small, we simply ran consecutive pairs of full simulations, re-computing the underlying network solution from the first of each pair with updated coincidence factors and loss times to correctly cost and dimension the network in the second. This is then repeated until convergence, with a check for possible oscillation. At a later date, the time series analyses will be fully embedded at critical stages of the planning algorithm, to prevent the need to run the full algorithm several times. Oscillation due to the decoupled treatment of the time series data must also be resolved.

III. Results

The network simulations are loosely based on a Finnish west-coast suburban/rural MV (20kV) network area, with 2 primary (110/20kV) substations and 74 secondary (20/0.4kV) substations. The outputs of the wind turbines are geographically modelled, using the methodology presented in [7], [8].

Fig. 4 shows the deterministically planned initial network, generated with coincidence factors of 1, i.e. assuming that the 95 percentile maximum or minimum (generation) demands occur simultaneously. Fig. 5 shows the final solution network, which corresponds to the coincidence values shown in the Appendix (calculated using the Monte Carlo simulation method). It is, of course, impossible to show the time series data used in this paper, but the time series data was created from appropriate proportions of 4 consumer types, district heating, offices and small business, direct electric heating, and electric storage heating, where the consumption is coupled with MV connected wind turbines at 12 locations (the wind generation is modelled using the Monte Carlo methods).

Table I summarizes the cost components of the solution networks. Note that two cost summaries are given for the initial network in Fig. 4. The first is for the assumption of unity coincidence factors and the second uses the utilization times for losses and line specific coincident factors estimated using the Monte Carlo method (re-computed for the same network topology). The results are as expected. Introducing the estimated coincidence factors for this lightly loaded network is more realistic, as it eliminates the overestimation of the losses. The solution network minimises interruption and investment costs at the expense of loss costs.

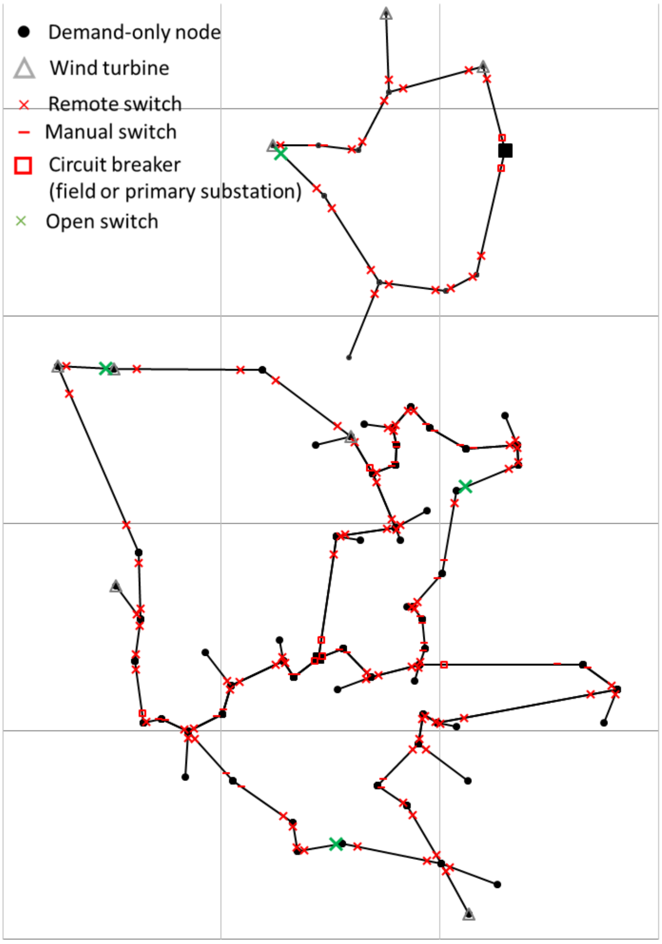

Figure 4 Initial deterministically planned network with symbol key, total costs are 5 545 312 €

The final solution network (Fig. 5) shows some component and topological differences to the network shown in Fig. 4, and is 3% cheaper than the deterministically planned network costed with power flows moderated by time-series derived coincident factors (this is considered the fairest comparison).

TABLE I. NETWORK SOLUTION COST COMPONENTS AND NO. OF PROTECTION DEVICES

Component	Initial network with coincident factors = 1	Initial network with time-series based loss times and coincidence factors	Solution network
Total cost	**5 545 312**	**5 425 059**	**5 266 970**
Total investment costs:	4 389 077	4 291 187	4 175 215
Total loss costs:	386 879	344 149	363 105
Total interruption costs:	769 355	789 723	728 649
No of manual switches:	30	33	37
No of master stations for remote switches:	39	38	36
No of remote switches:	97	93	88
No. of master stations for network circuit breakers:	3	3	4
No. of network circuit breakers:	3	4	5

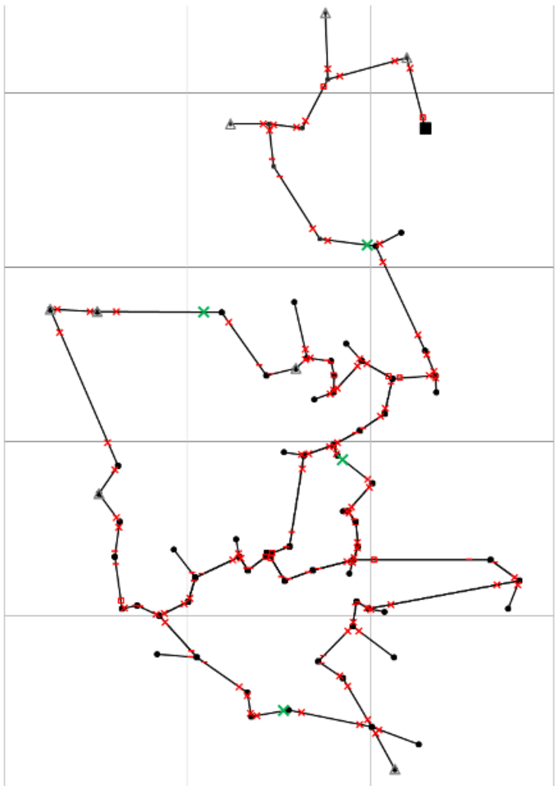

Figure 5 The final solution network using line specific coincidence factors (95th percentile) and loss times, total costs are 5 266 970 €

Note that the wind turbines are assumed to have their own protection devices. What these solution networks do not consider, is the potential for islanding parts of the network during faults upstream from distributed generation locations. Only the wind turbine locations behave as prosumer nodes, but in fact every node in the network is modelled as a prosumer, and can contain any mix of generation and consumption, as long as time series for each component can be provided.

Although the simulations allowed choosing between underground cable and bare overhead conductors, underground cables were mostly chosen by the planning algorithm, with only a few covered conductor overhead line sections in the solution networks.

IV. Discussion

The paper has made some significant steps towards planning contemporary and future distribution networks in our particular planning algorithm, and hopefully has offered some methodology of more general benefit. While the basic planning algorithm utilised in this paper includes a comprehensive treatment of conventional radial top-down reliability, optimally placing up to three layers of switching (circuit breakers, remote/automatic and manual load-breaking switches) [16], it does not yet do full justice to the supply possibilities of distributed generation, in terms of islanded operation during major grid contingencies. There should, for example, be the possibility of downstream circuit breakers that can isolate faulted upstream network (upstream and downstream referring to normal radial operation with no DG). The statistical modelling of consumption presented in the literature is quite advanced; however, the full ability of that modelling to generate thousands of Monte-Carlo simulated years has not been utilised in this paper. We chose to take the expected yearly profiles for a range of customer types, and superimpose them on one hundred years of simulated hourly wind data for 12 wind generation locations. The statistical modelling of limited customer bases is another challenge that we are addressing. Rural Finland, for example, has secondary substations that have only single customers behind them. The use of 95^{th} percentile confidence limits for the coincidence factors that scale down the load flows implied by using the 95^{th} percentile maximum and minimum net demand (negative for net generation) from the time series data in the line sections should mean that higher than 95% confidence is achieved in the planning process (as at the same time both the individual nodal maximums are calculated with 95 % CIs and the likelihood of them occurring at the same time is calculated with 95 % CIs).

A probabilistic approach may mean, on rare occasions, demand or distributed generation curtailing, and require a closer monitoring of the state of the network. A state estimation algorithm has been developed by the 4^{th} author [23].

International working groups [24] have been indicating that not only the stochasticity and divergence of distributed generation and consumption should be taken account of in the networks of the future, but also active network operation. This paper has developed a method to handle these challenges, provided they can be represented in simulated time series that cover the behaviour of all the constituent components in future networks.

The network plans shown in this paper show some benefit in terms of total costs compared to planning using conventional top-down logic. That there are at least marginal savings in total societal costs is perhaps self-evident, but we hope that clear methodology has been established showing one way to treat a complex problem. If we had focused on the expansion and upgrade of a more highly loaded network, the savings and topological changes in the network solutions would certainly have been greater than in the relatively lightly loaded Greenfield planning problem presented in this paper. It is to be admitted that the planning philosophy illustrated in this paper is still on the conservative side, but it may be argued that such long-term target networks should err on the conservative side, given that the future is hard to predict.

V. Conclusion

This paper presented a decoupled methodology suitable for planning and evaluating distribution networks with distributed generation, provided that full time series data and modelling is available for what lies behind the secondary substations. The probabilistic treatment and periodic processing of such time series data was coupled with a distribution network planning algorithm. The particular network section that we simulated showed some topological differences between the deterministic and probabilistic planning philosophies, but only marginal savings.

The methodology is an incremental development of that given in [13], which based nodal time series modelling on expected values for consumption and generation of a single *typical* year. We have now introduced the methodology to utilise full multi-year Monte-Carlo simulations of a properly modelled year; in this paper only demonstrated for wind generation, but in principle applicable to multiple consumption and generation types.

Acknowledgment

The authors would like to thank Helen Sähköverkko Oy, Loiste Sähköverkko Oy, and Trimble for their past, present and future cooperation. Aalto Energy Efficiency Research Programme STEEM and SAGA projects and the FLEXe project in Finland.

References

[1] M. Espinoza, C. Joye, R. Belmans, and B. De Moor, “Short-Term Load Forecasting, Profile Identification, and Customer Segmentation: A methodology based on periodic time series”, *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1622-1630, Aug. 2005.

[2] A. Mutanen, M. Ruska, S. Repo, and P. Järventausta, “Customer classification and load profiling method for distribution systems”, *IEEE Trans. Power Del.*, vol. 26, no. 3, pp. 1755–1763, Jul. 2001.

[3] F. A. Razak, A. H. Hashim, I. Z. Abidin, and M. Shitan, “Moving holidays’ effects on the Malaysian peak daily load”, *IEEE 2010 International Conference on Power and Energy*.

[4] A. S. Khwaja, M. Naeem, A. Anpalagan, A. Venetsanopoulos, and B. Venkatesh, “Improved short-term load forecasting using bagged neural networks”, *Electric Power Systems Research*, vol. 125, August 2015, pp. 109-115.

[5] M. Koivisto, P. Heine, I. Mellin, and M. Lehtonen, “Clustering of connection points and load modeling in distribution systems”, *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1255–1265, May 2013.

[6] M. Koivisto, M. Degefa, M. Ali, J. Ekström, J. Millar, and M. Lehtonen, “Statistical modeling of aggregated electricity consumption and distributed wind generation in distribution systems using AMR data”, *Electric Power Systems Research*, vol. 129, pp. 217–226, to be published in December 2015.

[7] M. Koivisto, J. Ekström, J. Seppänen, I. Mellin, E. Saarijärvi, and L. Haarla, "Statistical analysis of large scale wind power generation using Monte Carlo simulations", *Power Systems Computation Conference (PSCC) 2014, Wroclaw, Poland*, pp. 1–7

[8] J. Ekström, M. Koivisto, I. Mellin, J. Millar, E. Saarijärvi, and L. Haarla, "Assessment of large scale wind power generation with new generation locations without measurement data", Renewable Energy, vol. 83, pp. 362-374, to be published in November 2015

[9] A. Heidari, V. G. Agelidis, and M. Kia, "Considerations of sectionalizing switches in distribution networks with distributed generation," in *Power Delivery, IEEE Transactions on*, vol.30, no.3, pp.1401-1409, June 2015

[10] E. Saarijärvi, R.J. Millar, J. Niskanen, T. Nieminen, and M. Lehtonen, “Distribution network topology planning using life cycle cost driven cost surfaces in the internodal parameter computation,” *International Review of Electrical Engineering*, vol. 8, no. 1, pp. 395-402, Feb 2013

[11] E. Saarijärvi, M. Koivisto, J. Millar, M. Lehtonen, and J. Niskanen, "Generating fault rate surfaces using network fault statistics and geographic information," in *PES General Meeting*, *Conference & Exposition, 2014 IEEE* , 27-31 July 2014, pp.1-5

[12] R. J. Millar, S. Kazemi, M. Lehtonen, and E. Saarijärvi, "Impact of MV connected microgrids on MV distribution planning," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 2100-2108, Dec. 2012.

[13] R. J. Millar, E. Saarijärvi, M. Lehtonen, M. Degefa, and M. Koivisto, "Impact of node specific load growth and microgrids on distribution network planning, " *Cigré Paris Session 2014*, pp. 1-6

[14] R. J. Millar, E. Saarijärvi, M. Lehtonen, M. Hyvärinen, J. Niskanen, and P. Hämäläinen, “Electricity distribution network planning algorithm based on efficient initial and radial-to-full network conversion,” *International Review of Electrical Engineering*, vol. 8, no. 2, pp., Jun 2013

[15] R. Millar, E. Saarijärvi, and M. Lehtonen, An improved initial network for distribution network planning algorithms, (2014) *International Review of Electrical Engineering (IREE), 9*(3), pp. 538-547

[16] R.J. Millar, M. Lehtonen, and E. Saarijärvi, “Switch and reserve connection placement in a distribution network planning algorithm,” in *International Conference on Power System Technology* (*PowerCon*), 2012, pp. 1-6

[17] F. Pilo, G. Celli, S. Mocci, and G. G. Soma, "Multi-objective programming for optimal DG integration in active distribution systems," in *Proc. 2010 IEEE Power & Energy Society General Meeting*, pp.1-7

[18] S. Haffner, L. F. A. Pereira, L. A. Pereira, and L. S. Barreto, "Multistage model for distribution expansion planning with distributed generation – part II: numerical results," *IEEE Trans. Power Delivery,* vol. 23, April 2008, pp. 924-929

[19] B. Tammelin, T. Vihma, E. Atlaskin, J. Badger, C. Fortelius, H. Gregow, M. Horttanainen, R. Hyvönen, J. Kilpinen, J. Latikka, K. Ljungberg, N. G. Mortensen, S. Niemelä, K. Ruosteenoja, K. Salonen, I. Suomi and A. Venäläinen: “Production of the Finnish Wind Atlas,” *Wind Energy*, 2013, pp. 19–35

[20] www.winwind.com/WWD-1.aspx

[21] http://en.wind-turbine-models.com/turbines/732-enercon-e-12

[22] M. Z. Degefa, M. Koivisto, R. J. Millar, and M. Lehtonen, "Dynamic Thermal State Forecasting of Distribution Network Components," in *Probabilistic Methods Applied to Power Systems (PMAPS), 2014 International Conference on*, pp. 1-6.

[23] M.Z. Degefa, M. Humayun, A. Safdarian, M. Koivisto, R.J. Millar, and M. Lehtonen, “Unlocking distribution network capacity through real-time thermal rating for high penetration of DGs,” in *Electric Power Systems Research*, vol. 117, pp. 38-46, Dec. 2014

[24] F. Silvestro, A. Baitch, F. Pilo, B.B. Jensen, F. Mingtian G. Pisano, P. Georgilakis, N. Hatziargyriou, G. Petretto, "Demand side integration aspects in active distribution planning," in *Electricity Distribution (CIRED 2013), 22nd International Conference and Exhibition on*, vol., no., pp.1-4, 10-13 June 2013.

APPENDIX – NETWORK DATA

Space does not allow presenting all the network nodal data used in the simulations that produced Figs. 4 and 5, but Table II gives a few sample nodal parameters. The last column gives coincident factors associated with the line sections with the same reference number from the final solution network in Fig. 5.

TABLE II. NETWORK NODAL DATA SAMPLES

Node /line sect. ref. *n*	X-coord (km)	Y-coord (km)	P_{max} (kW)	P_{min} (kW)	Q_{max} (kW)	Q_{min} (kW)	T_{node_losses} (h/year)	$f_{coinc,95th}$ (p.u.)
0	0	0	0	0	0	0	0	0
1	-4.28	-12.24	0	0	0	0	0	0
2	-0.5	2	800	-1000	74	-75	1708	0.85
3	1.77	-12.4	431	0	107	0	2169	0.84
4	2.56	-13	492	0	122	0	1227	0.87
5	2.25	-13.8	492	0	122	0	2033	1
6	-4.34	-7.1	504	0	125	0	1227	1
7	-3.53	-6.9	80	-30	11	-10	2023	0.94
8	-3.86	-9.3	672	0	166	0	1227	0.86
9	-2.17	-6.2	611	0	151	0	1227	0.87
10	-1.73	-6.7	599	0	148	0	2033	0.86
11	-2.58	-6.7	800	0	198	0	1227	0.88
12	-3.23	-6.6	788	0	195	0	510	1
13	0	-6.4	547	0	135	0	1227	1
14	-0.89	-7.2	438	0	108	0	1227	0.85
15	0.28	-7.1	718	0	178	0	2169	0.84
16	0.3	-7.6	718	0	178	0	1227	0.94
Missing data								
48	-6.23	-15.2	390	0	97	0	1227	0.84
49	-1.98	-14.3	603	0	149	0	2169	0.83
50	-0.86	-15.2	310	0	77	0	2033	1
51	-2.93	-15.3	495	0	122	0	510	1
52	-2.24	-15.8	495	0	122	0	510	1
53	-2.73	3.3	1025	-1000	125	-120	1078	1
54	-3.37	0	470	0	116	0	2033	0.99
55	-4.27	0.1	470	0	116	0	1227	1
56	-5.31	0.1	996	-1000	112	-100	1276	1
57	-2.66	1.4	1569	0	59	0	510	0.93
58	-0.65	-3	514	0	127	0	2169	0.81
59	-1.36	-3.4	514	0	127	0	2169	0.79
60	-2.89	-3.2	246	0	61	0	2033	0.77
61	-3.58	-5	347	0	86	0	2033	0.82
62	-4.15	-1.1	350	0	86	0	2033	1
63	-5.55	-5.3	2550	0	35	0	2169	0.86
64	-8.39	-9.7	446	0	110	0	2169	0.77
65	-8.35	-11.3	436	0	108	0	1227	0.82
66	-8.92	-10.5	70	-30	8	-8	2006	0.79
67	-8.48	-12.3	425	0	105	0	2033	0.84
68	-3.72	-16.7	408	0	101	0	2169	0.71
69	-4.87	-16.2	572	0	141	0	2033	0.8
70	-4.75	-16.9	609	0	151	0	2033	0.76
71	-1.46	-17.2	301	0	74	0	510	1
72	-0.19	-17.7	450	0	111	0	510	1
73	-0.83	-18.4	461	-30	0	0	501	1
74	-8.95	-5.28	461	-1000	0	0	1736	0
75	-10.25	-5.2	461	-30	0	0	1208	1

The cost and electro-technical parameters of the underground cables and covered overhead conductors are given in Tables III and IV.

TABLE III. UNDERGROUND CABLE PARAMETERS

Fixed costs (k€/km)	Resistance (Ω/km)	Reactance (Ω/km)	Susceptance (μS/km)	Thermal Limit (A)
30.44	1.20000	0.16022	40.8	110
33.16	0.64100	0.14451	50.3	155
43.86	0.38000	0.12881	66.0	235
46.31	0.25000	0.12252	75.4	300
52.76	0.15000	0.10996	94.2	385

TABLE IV. COVERED OVERHEAD CONDUCTOR PARAMETERS

Fixed costs (k€/km)	Resistance (Ω/km)	Reactance (Ω/km)	Susceptance (μS/km)	Thermal Limit (A)
27.990	0.493	0.302	3.770	310
31.930	0.288	0.284	4.084	430
37.052	0.188	0.27	4.398	560
73.864	0.094	0.135	8.796	1120

Note that, for convenience, interruption costs were set to 1.1€/kW/fault and 11€/kWh for all nodes, which are rather high values for rural customers. The planning horizon is 40 years, the annual interest rate is 6 % and net annual nodal load growth is set globally to 0.12 %. These parameters can be node specific if required.

R. John Millar received a Bachelor of Engineering Degree (mechanical) from the University of Auckland, New Zealand, in 1984. He completed a Master's Degree in Electrical Engineering in 2002 and a D.Sc. (Tech) in 2006 at the Helsinki University of Technology, Finland. This university is now part of Aalto University, where John continues to work as a researcher and lecturer. His main research interests are distribution network planning and underground cable rating.

Jussi Ekström was born in Finland in 1988. He received the Bachelor of Science (Technology) and the Master of Science (Technology) degrees in electrical engineering in 2012 and 2014, respectively, from the School of Electrical Engineering, Aalto University, Espoo, Finland, where he is currently working toward the Doctoral degree. He is also a Researcher with the School of Electrical Engineering, Aalto University. His main research interests include applying statistical and machine learning methods to the modeling of wind and solar energy generation and their impact on the power system.

Matti Lehtonen (M'11) was with VTT Energy, Espoo, Finland from 1987 to 2003, and since 1999 has been a professor at the Helsinki University of Technology, nowadays Aalto University, where he is now head of Power Systems and High Voltage Engineering. He received both his Master's and Licentiate degrees in Electrical Engineering from Helsinki University of Technology, in 1984 and 1989 respectively, and the Doctor of Technology degree from Tampere University of Technology in 1992. His main activities include power system planning and asset management, power system protection including earth fault problems, harmonic related issues and applications of information technology in distribution systems.

Matti Koivisto is working as a Postdoc at the Technical University of Denmark, Department of Wind Energy. His research focuses on stochastic modelling of large-scale wind and solar generation, and studying their effects on the Nordic and European power systems. He is also experienced in electricity consumption modelling. He has Doctor of Science (Technology) and Master of Science (includes the Bachelor of Science) degrees from Aalto University School of Electrical Engineering from 2016 and 2010, respectively.

Eero Saarijärvi (1980) received the M.Sc. degree in Electrical Engineering from Helsinki University of Technology in 2008 and a D.Sc. (Tech.) in 2014 from the same university, now called Aalto University. His doctoral studies concentrated on the topographical and topological considerations of Finnish medium voltage distribution networks. After working as post-doc researcher and lecturer at Aalto University, Eero is now working for Trimble.

Merkebu Z. Degefa received a B.Sc. in electrical engineering from Mekelle University, Ethiopia, in 2007, an M.Sc. with distinction from Aalto University, Finland, in 2010, and a D.Sc. (Tech.) from Aalto in 2015. After a period of post-doc research at Aalto, Merkebu has taken up a position as a research scientist at SINTEF Energi AS in Trondheim, Norway. His main research interests are active distribution system management, dynamic thermal rating, demand response, and load modeling.