CAPACITY EXPANSION MODELLING

PREPARED BY:	Supply Forecasting
DOCUMENT REF:	SF-PC-CAPMDL
VERSION:	2.0
DATE:	28 November 2014

DRAFT

Contents

1 Introduction

This document describes AEMO's electricity capacity expansion modelling. The description is technical in nature, and is intended to be read in conjunction with AEMO's planning consultation methodology and input assumptions report[1].

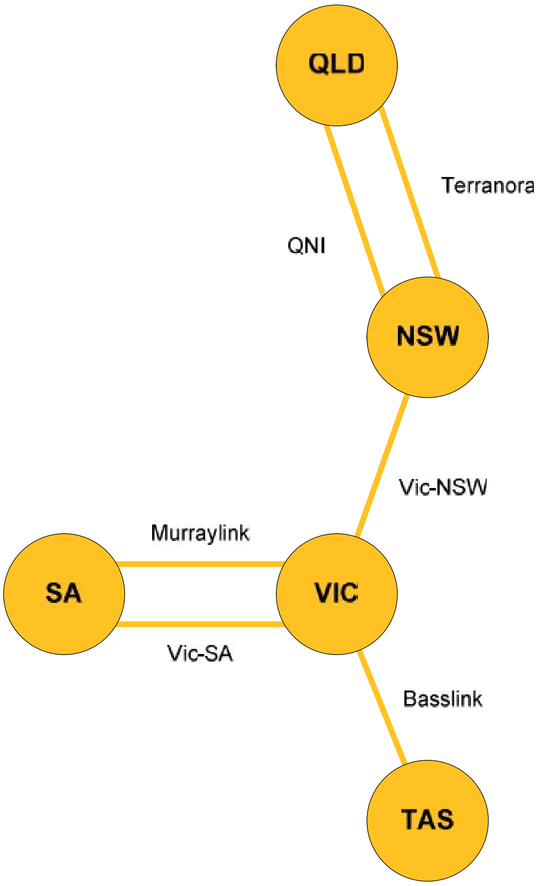
Capacity expansion modelling is the process of optimally adding new generation and transmission to the NEM as future demand grows. The set of new generation and transmission facilities and the time at which they are activated is referred to as the *expansion plan*. In this case, "optimal" refers to the expansion plan that results in the lowest combined operating, capital and reliability cost incurred by operating the NEM securely over the course of the modelled time horizon. Expansion plans are used in a range of studies, including the National Transmission Network Development Plan (NTNDP), feasibility and pre-feasibility studies, and some Regulatory Investment Test – Transmission (RIT-T) studies that AEMO undertakes in its role as planner of the Victorian Declared Shared Network (DSN).

AEMO is continually developing a capacity expansion model that selects generation and transmission projects to include in the future network from a series of generation technology, fuel supply and transmission project options. It is implemented using PLEXOS for Power Systems from Energy Exemplar.

2 Topology

The model uses a regional representation of the NEM. Each NEM region is implemented as a node, connected radially by interconnectors as shown in Figure 1. This topology reflects the topology used operationally by the NEM Dispatch Engine (NEMDE).

Figure 1 - Regional representation of the NEM



[1] Available from http://www.aemo.com.au/Consultations/National-Electricity-Market/~/media/Files/Other/planning/2013Consultation/Planning_Studies_2013_Methodology_and_Input_Assumptions.ashx

3 Dispatch cost

Electricity demand and generation occurs at the nodes of the model. Power flows between nodes on notional interconnectors. Electricity prices are developed by the model at each node. The model considers a collection of time intervals of varying length, developing a price in each time interval. By multiplying generation cost and demand and summing the product over all regions and all time intervals, a cost of dispatch is developed:

$$C_d = \sum_{t=1}^{T} \sum_{g=1}^{GA} CGA_{t,g} + \sum_{g=1}^{G} CG_{t,g} * E_{t,g} + \sum_{r=1}^{R} CP_{t,r} * USE_{t,r} \qquad \text{Eq. 1}$$

where

C_d is the cost of dispatch

T is the number of time intervals considered

R is the number of regions

GA is the number of available generators (incurring fixed operating costs)

CGA is the cost to be available at time t for generator g

G is the number of generators

$CG_{t,g}$ is the cost to generate at time t for generator g

$E_{t,g}$ is the energy supplied by generator g at time t

$CP_{t,r}$ is the penalty cost for not supplying energy at time t in region r

$USE_{t,r}$ is the energy not supplied at time t in region r

Due to the way in which demand is treated in the model, unserved energy (USE) is not allowed to occur in dispatch and the second term in equation 1 is zero. Instead, minimum reserve levels and growing demand place a lower bound on the amount of generation that must be present to ensure a secure supply. The time-sequential model, which has much higher time resolution, is used to ensure capacity expansion outcomes conform to the Reliability Standard.

Generation cost is the sum of:

- Fixed operating cost, independent of generation output and incurred as long as a generating unit is available (CGA in equation 1)
- Variable operational cost, dependent on generation output
- Fuel cost, determined by multiplying generation output by a heat rate for thermal generators to calculate fuel usage, then multiplied by the generator's fuel price
- Emissions cost, determined by multiplying the generation output by the unit's CO2-e emissions intensity factor to calculate total emissions, then multiplied by emission price.

Energy generated by a generating unit is the sum of the generator's share of:

- Energy to supply demand
- Energy lost in transportation within a region, an assumed component of demand
- Energy lost in transportation between regions, controlled by interconnector loss equations
- Energy to supply the generating unit itself, called the auxiliary load

3.1 Price and cost

The optimisation minimises the cost of energy production and transport. In every time interval considered, the most expensive generating unit, called the marginal unit, sets the price of energy in each region. Theoretically, the true cost to customers in the NEM is the product of the amount of energy and the price of that energy in each region. Every generating unit with costs lower than the marginal unit should be able earn a profit, because generators pay according to their costs, but earn according to the price.

In practice, the model does not have the fidelity in time for price information to be useful in determining generator profitability. AEMO uses its time-sequential model to validate capacity expansion modelling outcomes and ensure that generator revenues fall into a reasonable range.

4 Generation expansion

In some time intervals, there may be insufficient generation to meet demand and requirements for minimum reserve, and when that is the case some demand (or reserve) cannot be supplied. The model includes representations of new generation that may be established to prevent the occurrence of supply shortfall. Using these, the model develops a cost of new generation:

$$C_{ng} = \sum_{ng=1}^{NG} \sum_{t=1}^{T} CA_{t,ng} \qquad \text{Eq. 2}$$

where

C_{ng} is the cost to establish new generation

T is the number of time periods considered

NG is the number of new generating unit options

$CA_{t,ng}$ is the amortised capital cost of generation, which varies depending on the size of the time interval under consideration

The cost of establishing new generation is balanced by the operational cost benefit provided by generation sources that are more efficient or have lower fuel, variable, fixed or carbon costs, allowing the model to select the option with the lowest total cost that is appropriate at a specific location and time. The cost of adding new generation to the model solution is amortised (spread over the modelled timeframe) to allow the model to choose an appropriate time to activate it. The model is able to choose from more than 170 new generating unit options, which are permutations of technology type and location. The model allows for new generation expansion from 2017-18. AEMO considers that new projects that are not committed are likely to require a minimum of 18 months to two years be fully commissioned and generating in the market.

4.1 Binary and continuous decision-making

In general, there are two ways of adding new generation to the network: either as individual generating units of specific size, or as a generic generation capacity that is allowed to vary continuously. The former method is more intuitive and allows costs to be estimated with higher confidence (the cost of a 150MW open cycle gas turbine, for example, is well-known), but results in a “mixed integer program” that quickly becomes computationally intractable as more options are added. The latter method results in a problem with much lower computational overhead, but tends to result in unfamiliar new generating unit capacities and costs that are more difficult to confirm (the cost of a 94.2MW open cycle gas turbine, for example, is unlikely to be supported by information gleaned from actual projects, but could be estimated by interpolating the cost of a 70 MW unit and a 150 MW unit).

To ensure manageable computation time, AEMO uses the continuous generation capacity method.

Like existing generation, new generation incurs operational costs that are added to the dispatch cost equation (Eq. 1).

5 Generation retirement

The capacity expansion model allows for existing generators to retire if simulated market signals indicate that costs may be minimised by replacing existing capacity with new capacity with a combination of lower fuel, emission or operating costs, or a location proximal to demand (reducing costs due to losses).

Generation may also be withdrawn from participation in the model if a generator has advised AEMO of its intention to decommission generating capacity.

The cost of retiring existing generation is balanced by the operational cost benefit provided by generation sources that are more efficient or have lower fuel, variable, fixed or carbon costs, allowing the model to select the option with the lowest total cost that is appropriate at a specific location and time.

In 2014, the model tested for plant retirements beyond 2017-18. No retirements are allowed in the first 2 years of the simulation as it's considered unlikely for plant to retire from the market without any prior market notification within this period.

The model is able to retire from all existing coal and gas fuel generators that have reached 80% of technical life. Unlike the continuous generation capacity method used in the expansion plan, the retirement decisions are made on the total existing plant capacity.

Cost of plant retirement includes the cost of end of life plant remediation and site rehabilitation as obtained from the fuel and technology cost review by ACIL Allen in 2014. These costs are technology specific.

$$C_{rg} = \quad Unit\ Retierement\ cost \qquad \text{Eq. 3}$$

C_{rg} = Total cost of plant retirements.

6 Transmission augmentation

In some cases, it may be more cost-effective to build generation in a location far from centres of demand and augment the transmission network in order to provide access to that generation. This can be the case where an excellent renewable resource exists some distance from the existing network, or when low fuel prices encourage the development of a generation cluster in a particular area.

The model contains a representation of new interconnector projects that can increase power transfer between regions. The increased power transfer can reduce costs by providing access to cheaper, more remote generation, generation that is active at a different time (for example, solar generation in South Australia peaks later in the day from the perspective of customers in Victoria, compared to solar generation in Victoria) or by reducing transmission losses. Transmission projects introduce a new cost to the system in the same way as new generation projects:

$$C_{nt} = \sum_{nt=1}^{NT} \sum_{t=1}^{T} CA_{t,nt} \qquad \text{Eq. 4}$$

where

C_{nt} is the cost to establish new transmission

T is the number of time periods considered

NT is the number of new generating units available

$CA_{t,nt}$ is the amortised capital cost of new transmission, which varies depending on the size of the time interval under consideration

Unlike the generation expansion, transmission augmentations are modelled as yes/no decisions using integer variables. The computational cost of doing so is much smaller than that of using integer decisions for generation, because the number of options considered is much smaller.

7 Co-optimised expansion

The aim of the capacity expansion modelling is to minimise the total cost of meeting demand. This cost is given by the sum of equations 1 to 4 above (the sum of dispatch, new generation and new transmission costs):

$$minimise(C_d + C_{ng} + C_{nt} + C_{rg}) \qquad \text{Eq. 5}$$

An increase in the cost of establishing new generation or transmission is generally accompanied by a decrease in dispatch costs – either due to reduced load-shedding penalties, the improved efficiency of new generation, its lower operating or fuel costs or its better access to the changing load. The result is referred to as a generation and transmission co-optimised least cost expansion plan.

7.1 Energy and capacity

Equation 5 describes an optimisation in *capacity* and *energy*. It is driven by both:

- The demand for energy at peak times (maximum demand) and the need to provide capacity over and above maximum demand to ensure the power system continues to operate reliably even when some generation fails.
- The need to provide energy at all other times at the lowest possible cost

In any one time period, a region may source capacity and energy from its own generation or a neighbouring region. Minimum reserve levels specify a lower bound on the capacity that must be available at all times in each region (whether it comes from local generation or interconnectors, called *reserve sharing*). Costs are used to fine-tune generation output in each region, finding the lowest cost solution that does not violate the requirement to supply reserve.

The situation is complicated by intermittent generation, which cannot be reliably dispatched and so does not have a clearly specified capacity. For example, when adding up available capacity to determine how much minimum reserve to allow for, it is not clear how much capacity a wind farm can supply, because it is not clear how hard the wind will be blowing at that time. To manage this aspect of intermittent generation, the model is supplied with values that specify the portion of total installed intermittent generation capacity that may be contributed by that generation in any time interval. For wind generation, these "contribution factors" are specific to each region.

Separate day and night solar contribution factors are not specified, because the use of load blocks for specifying demand destroys the time-of-day information necessary to apply the factors rigorously.

8 Load blocks

AEMO's capacity expansion modelling may consider very long time periods of more than 20 years. The primary input to the model is the expected change in demand in each region. Operationally, NEMDE dispatches generation every 5 minutes (the dispatch interval), and sets prices every 30 minutes (the trading interval). The capacity expansion model must consider every time interval in its time horizon simultaneously, and using the NEM's dispatch or trading intervals leads to unreasonably large computational problems.[2] The time-sequential model uses hourly trading intervals, which are manageable because each hour is considered in isolation, but this is impractical for capacity expansion modelling.[3]

To make the capacity expansion problem tractable, the time-sequential model's projected hourly demand profile for each region is sorted and aggregated into "load blocks". This is illustrated in Figures 2, 3 and 4. Figure 2 shows a regional demand profile for a single year (in this case, New South Wales in calendar year 2012).

When the values in Figure 2 are sorted from highest to lowest, the load duration curve shown in Figure 3 is the result.

[2] There are 350,640 trading intervals in 20 years.

[3] The model is represented in the computer as a series of rows in a matrix. The time taken to calculate a solution increases approximately according to the square of the number of rows. A problem in a single time interval with n rows will require n^2 time, but the same problem considering two time intervals will require $(2n)^2 = 4n^2$ time. As an indication, if a time-sequential model could produce a solution for a single trading interval in 1 second, it would take $(350{,}640)^2$ seconds, or approximately 3,900 years to simultaneously solve for all trading intervals in a 20 year period.

Figure 2 - A regional demand time series

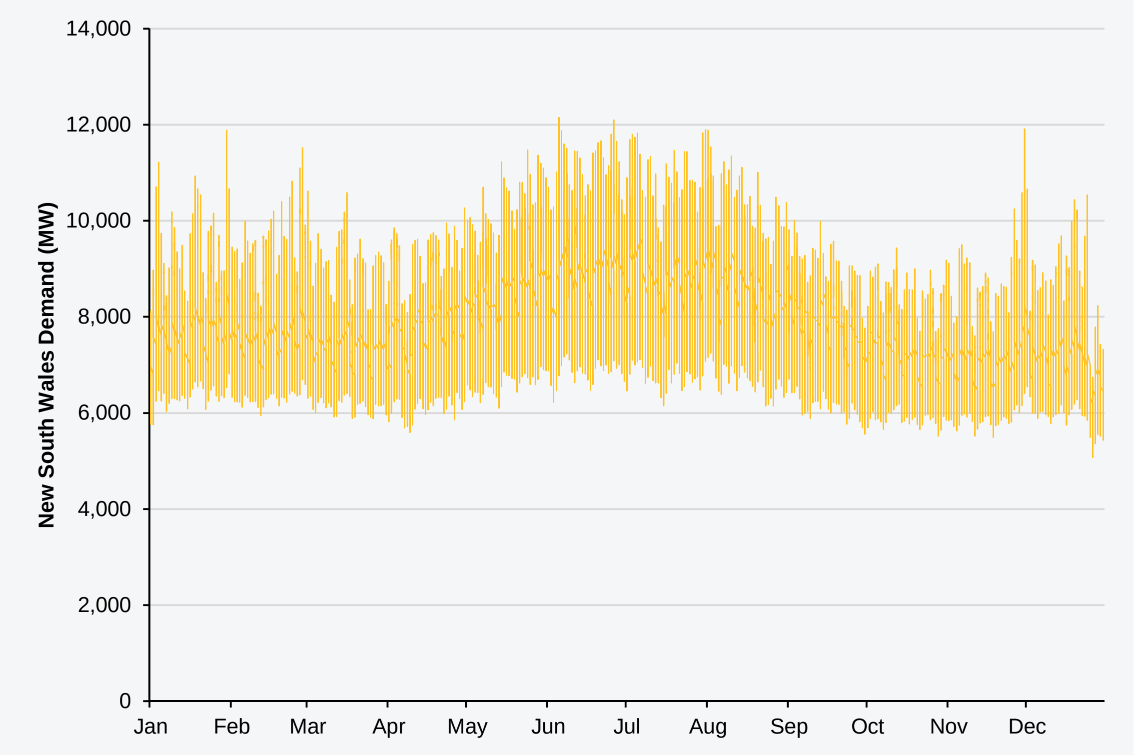
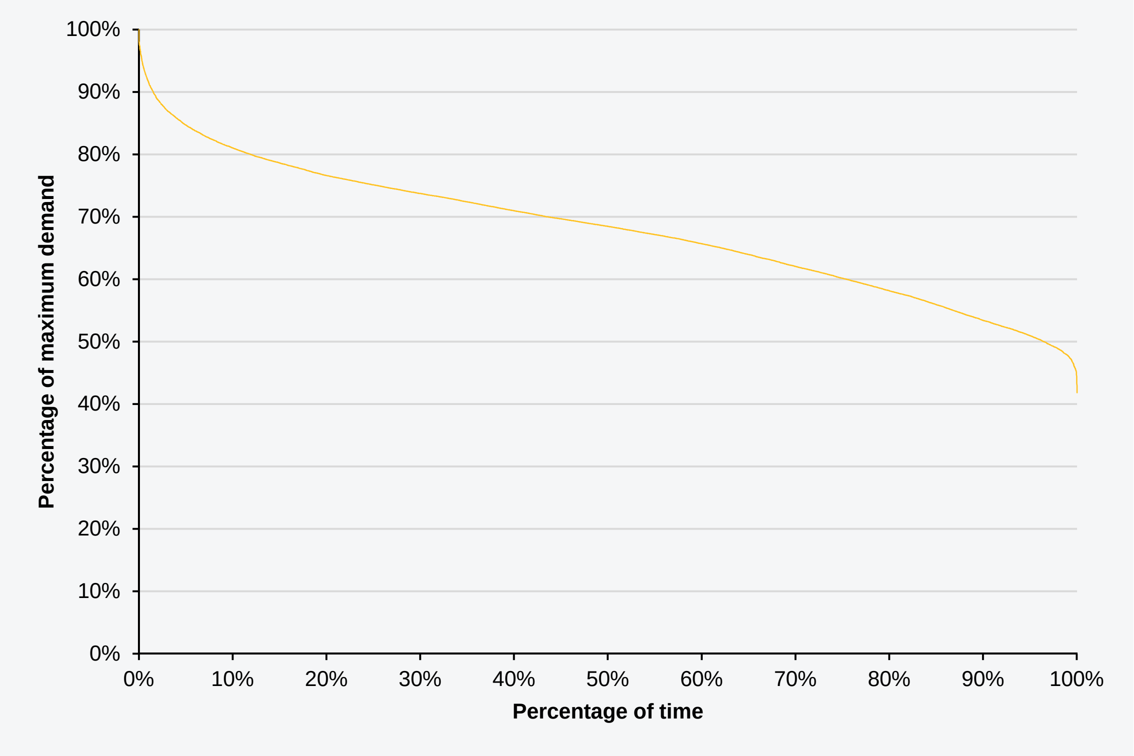
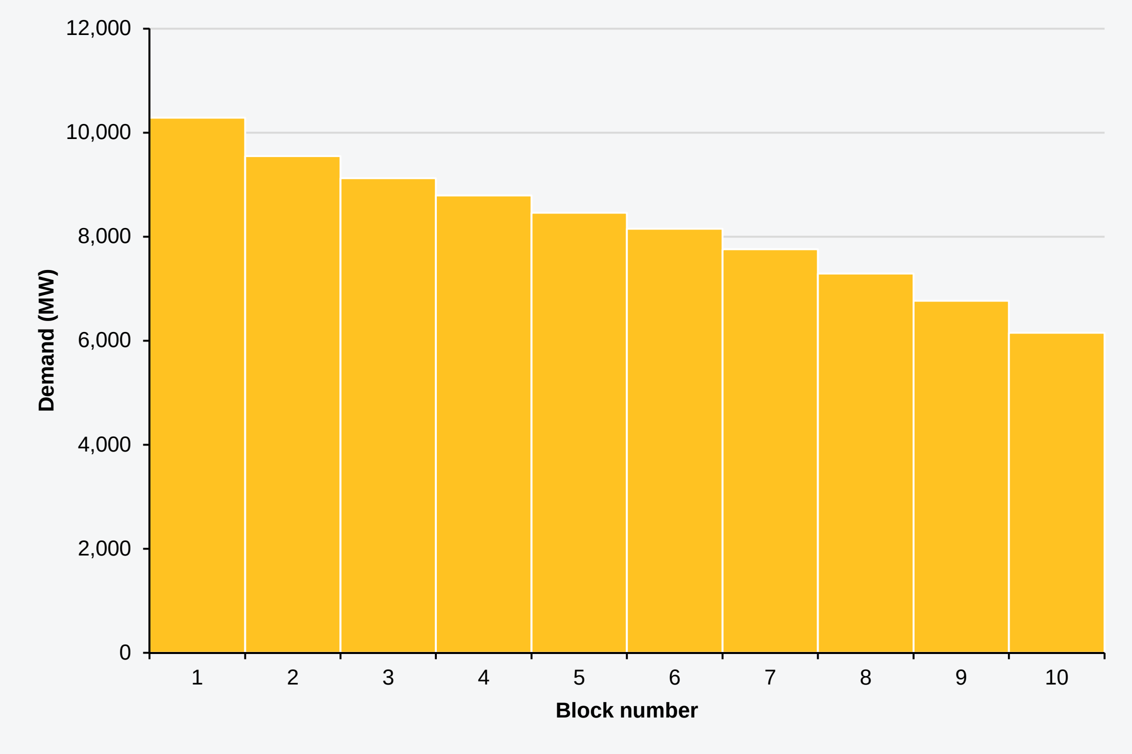


Figure 3 – Load duration curve of the time series in Figure 2



The load duration curve destroys the time relationship between values, but makes it easy to determine statistics like maximum, minimum and median demand at a glance, and develop knowledge about the impact of high demand times. For example, it is a simple matter to read off the chart that the top 20% of capacity is needed only 12% of the time, while the top 10% of capacity (about 1,200 MW, reserve levels notwithstanding) is needed only 1.5% of the time (or about 5.5 days in the year). It also suggests the concept of a "load block", whereby the load duration curve is split into chunks that represent a longer period of time. Figure 4 shows an example partitioning of the load duration curve in Figure 3. Here, the median load has been selected in each 10% interval of time.

Figure 4 - A 10-block version of the load duration curve in Figure 3



Using this method, the number of time periods that would need to be considered falls from 17,568[4] to 10.

There are many ways in which a load duration curve can be partitioned into load blocks. The method outlined above does not capture either the maximum or the minimum demand. The development of blocks of constant width ignores some the information available in the curvature and slope of the load duration curve. To circumvent these shortcomings, AEMO uses a weighted least-squares fit method to partition the load duration curve. With this method, the number of desired blocks is chosen, and the width of the blocks is optimised by the modelling software to minimise the deviation of the value of the block from the values in the load duration curve that the block will represent. In addition, the maximum demand interval is preserved as single-interval block.

AEMO uses between 12 and 20 blocks for each modelled month. When 20 blocks are chosen, the weighted least-squares fit method with preserved maximum results in 1 block of duration one hour to represent maximum demand and 19 blocks of varying duration that optimally follow the load duration curve.

8.1 Load diversity

Load diversity is a phenomenon whereby demand in one area may be low when demand in a neighbouring area is high. When this occurs frequently, it becomes apparent that transmission augmentation may be a more cost-effective solution than establishment of new generation, because surplus generation in the low demand region is available to supply the high demand region. Conversely, if two areas frequently experience high demand at the same time, power transfer between them is expected to be small and generation investment is likely to be more cost-effective because transmission investment may not necessarily translate into higher or more frequent power transfer.

Using load blocks to represent demand reduces the capacity expansion model's ability to make investment decisions based on load diversity, because it considers only the diversity present in the interval chosen to represent the load block. AEMO's time-sequential model, which considers the modelled horizon at hourly resolution, and power system models, which specifically target times of high demand, are used to verify the expansion plan outcomes.

[4] 2012 was a leap year - 366 days of 48 half-hour trading intervals each

9 Energy constraints

The time-sequential model contains a detailed representation of hydroelectric generators, their associated storage reservoirs and expected annual inflows into storage. The capacity expansion model cannot track the rise and fall of reservoir storage levels because it does not consider time continuously, and detailed accounting of water quantities and values is not possible.

No new hydroelectric capacity is included in the capacity expansion model, because it is assumed that all of the major water catchments that can be used for hydroelectric generation in the NEM have already been developed. In addition, hydroelectric inflows are assumed to follow long term averages, so hydroelectric generation does not fluctuate from year to year. Under these conditions, it is reasonable to assume hydroelectric generation output on an annual basis.

The capacity expansion model incorporates energy targets for hydroelectric generation. Hydroelectric generators may generate up to the target in each year. Generation beyond the target is possible, but a high penalty price is imposed on any generation over and above the target. In this way, the storage levels of the hydroelectric reservoirs are controlled to remain close to their current levels throughout the simulated horizon.

10 Transmission Limitations

The capacity expansion model does not incorporate a representation of intra-regional transmission limitations. AEMO uses the time-sequential and power system models to assess the impact of intra-regional transmission limitations on the expansion plans produced by the capacity expansion model. Where the time-sequential or power system model indicates that an expansion plan is materially impacted by intra-regional transmission limitations, either a special constraint equation is developed to limit generation expansion, or the costs of affected generation are modified to include the additional transmission investment required to raise the limitations, and the expansion plan re-developed.

The maximum allowable power flow on interconnectors is explicitly modelled, along with the maximum capacity reserve an interconnector can supply to energy-importing regions. Interconnector augmentation projects are defined in the model that can increase power transfer capability provided the cost of augmentation can be recovered through lower operating costs. These augmentations may originate from a range of sources, including jurisdictional planning body (JPB) annual planning reports (APRs), joint feasibility and pre-feasibility studies or augmentation studies undertaken internally at AEMO.

In 2014, several options were considered for the transmission interconnector upgrades. Details of these options are included in the additional modelling data[5].

11 Build limits

Every new generation technology is subject to a build limit, which defines how much new capacity of any one type can be installed in any one planning zone over the modelled horizon.

In some cases build limits may cross zone or technology boundaries, and special limit formulations are required. For example, limitations on gas supply may impose a build limit of 1,000 MW on new gas generation in a particular zone, which may receive either new combined-cycle gas turbines (CCGT) or open-cycle gas turbines (OCGT). Because of the way the model implements build limits, this fuel supply limitation implies three limits in the model:

- New CCGT generation may not exceed 1,000 MW
- New OCGT generation may not exceed 1,000 MW
- The sum of new CCGT and OCGT generation may not exceed 1,000 MW

[5] http://www.aemo.com.au/Electricity/Planning/Related-Information/~/media/Files/Other/planning/Planning_Studies_2013_Additional_Modelling_Data_July.ashx

When this is the case, special build limits may be defined for specific technologies, in a specific region, or for a particular fuel supply source. The generation technology expansion constraints are based on the ACIL Allen tech review in 2014[6].

New generation is built to meet the capacity and energy demand without any consideration of the plant development and planning timelines which can lead to substantial generation expansion in the early years of the expansion plan. This is modelled by limiting new generation to 1800 MW per year across the NEM prior to 2019-20.

12 Special constraints

12.1 Wind generation expansion in New South Wales

New wind farm generation is modelled using the wind capacity factor, marginal loss factor and cost of connection at the zone connection point. This results in the model planting substantial levels of new wind farm projects in New South Wales. However, considering the timelines for constructability of new projects AEMO has limited the wind generation expansion in the first eligible year of the expansion plan in 2018-19 to be 400 MW based on market information of publicly announced projects for New South Wales.

12.2 Generation expansion in South East Queensland

The South East Queensland zone is one of the smaller zones defined for generation and transmission expansion and is heavily populated. Its commercial and tourist industries are generally incompatible with heavy industry, and planning permission for new power stations is difficult to obtain. The zone has good access to load and gas, coal and solar resources, which lead the model to establish large amounts of generation there when alternative uses for land are not taken into account.

To manage expansion in the South East Queensland zone, a limit of 1,200 MW is imposed on all new generation there.

13 End effects

When a model includes amortised costs for new generation and transmission infrastructure, only a small portion of the total capital cost is apportioned to each load block. When the modelled time horizon is shorter than the expected lifetime of the new infrastructure assets, simply summing costs over the modelled horizon underestimates the true capital cost. This would bias investment towards the end of the modelled horizon, because later investments would effectively be much cheaper than earlier investments. Conversely, if all of the capital costs of an investment are apportioned to the single time period in which it occurs, there may not be sufficient time in the modelled horizon to recover the capital expenditure, biasing outcomes towards the beginning of the modelled horizon. This property of models is termed an “end effect” or “terminal effect”.

To manage end effects, models need some way of extending costs beyond the end of the modelled horizon to ensure that capital costs are recovered over the expected lifetime of the asset.[7] AEMO uses an “in perpetuity” method for this purpose. In perpetuity assumes that the costs calculated in the final year of the modelled horizon continue for all years beyond the horizon. A discount rate is applied to costs, which progressively reduces the impact of future costs on the present day value of investments until they become negligible.

AEMO uses a 10% discount rate for in-perpetuity calculations.

[6] http://www.aemo.com.au/Electricity/Planning/Related-Information/~/media/Files/Other/planning/2014%20Assumptions/Fuel_and_Technology_Cost_Review_Data_ACIL_Allen.ashx

[7] “Recovery” of capital costs has a special meaning in this context. The model must recover the money spent on a capital investment as a reduction in total operating costs across the NEM, not as an individual facility or company portfolio.

14 Related documents and data

Document	Description
Planning consultation assumptions and methodology	Describes AEMO's long term planning electricity and gas models and the assumptions behind their operation. Available from http://www.aemo.com.au/Electricity/Planning/Related-Information/~/media/Files/Other/consultations/nem/2014_Planning_Consultation_Methodology_and_Input_Assumptions_30_jan_14.ashx
New entry generation data	Cost and capability data for new generation that may form part of an expansion plan. Available from http://www.aemo.com.au/Electricity/Planning/Related-Information/~/media/Files/Other/planning/2014%20Assumptions/Fuel_and_Technology_Cost_Review_Data_ACIL_Allen.ashx
Minimum reserve levels	Used to set a lower bound on generation and interconnector import capacity available to meet regional demand. Available from http://www.aemo.com.au/Electricity/Planning/Related-Information/Assessing-Reserve-Adequacy
Additional modelling data	Data about economic conditions and physical limitations of transmission network elements. Used in the determination of least cost and the capability of the existing network. Available from http://www.aemo.com.au/Consultations/National-Electricity-Market/~/media/Files/Other/planning/2013Consultation/Planning_Studies_2013_Additional_Modelling_Data.ashx
Existing generation data	Data about the existing generation fleet. Available from http://www.aemo.com.au/Electricity/Planning/Related-Information/~/media/Files/Other/planning/Emissions%202014/Planning_Studies_2013_Existing_Generator_Technical_Data_230514.ashx