Sydney WATER

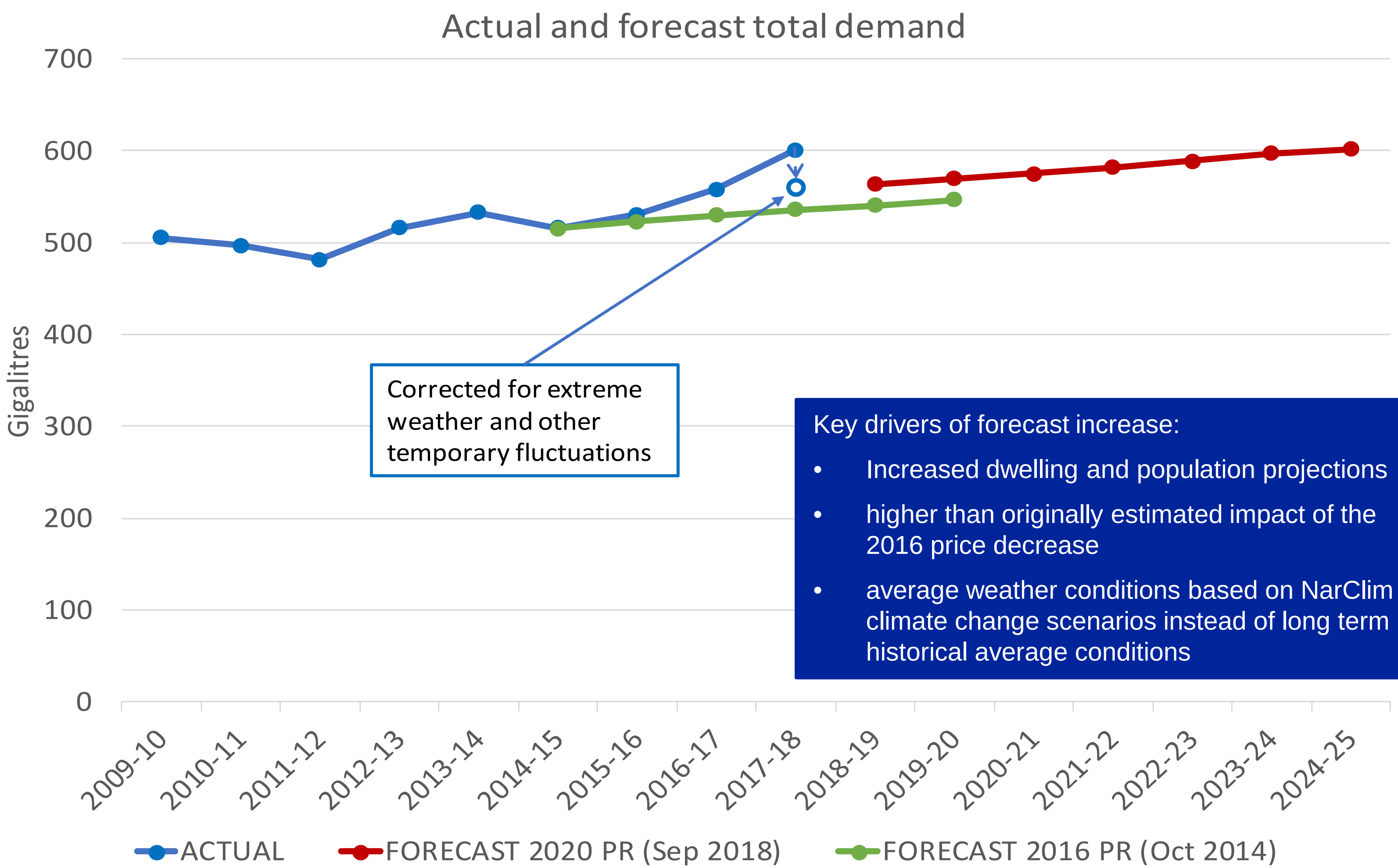
Demand forecast model for PR2020

Efficiency Review | August 2019

Contents

- The forecast and key assumptions
- Forecasting approach overview
- Residential model
- Non-residential model
- Billed unmetered and non-revenue water

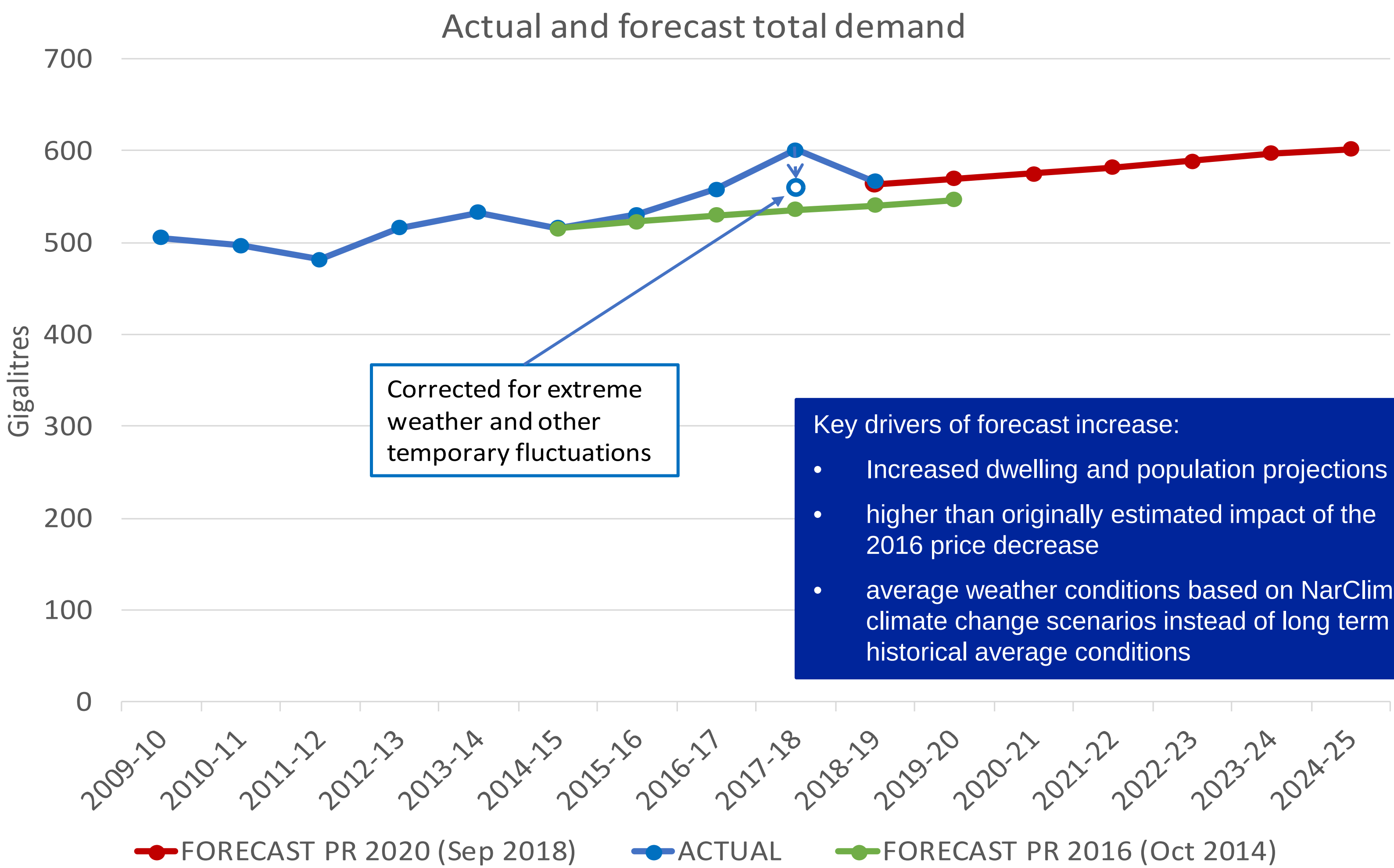
Total demand forecast



Key drivers of forecast increase:

- Increased dwelling and population projections
- higher than originally estimated impact of the 2016 price decrease
- average weather conditions based on NarClim climate change scenarios instead of long term historical average conditions

Total demand forecast

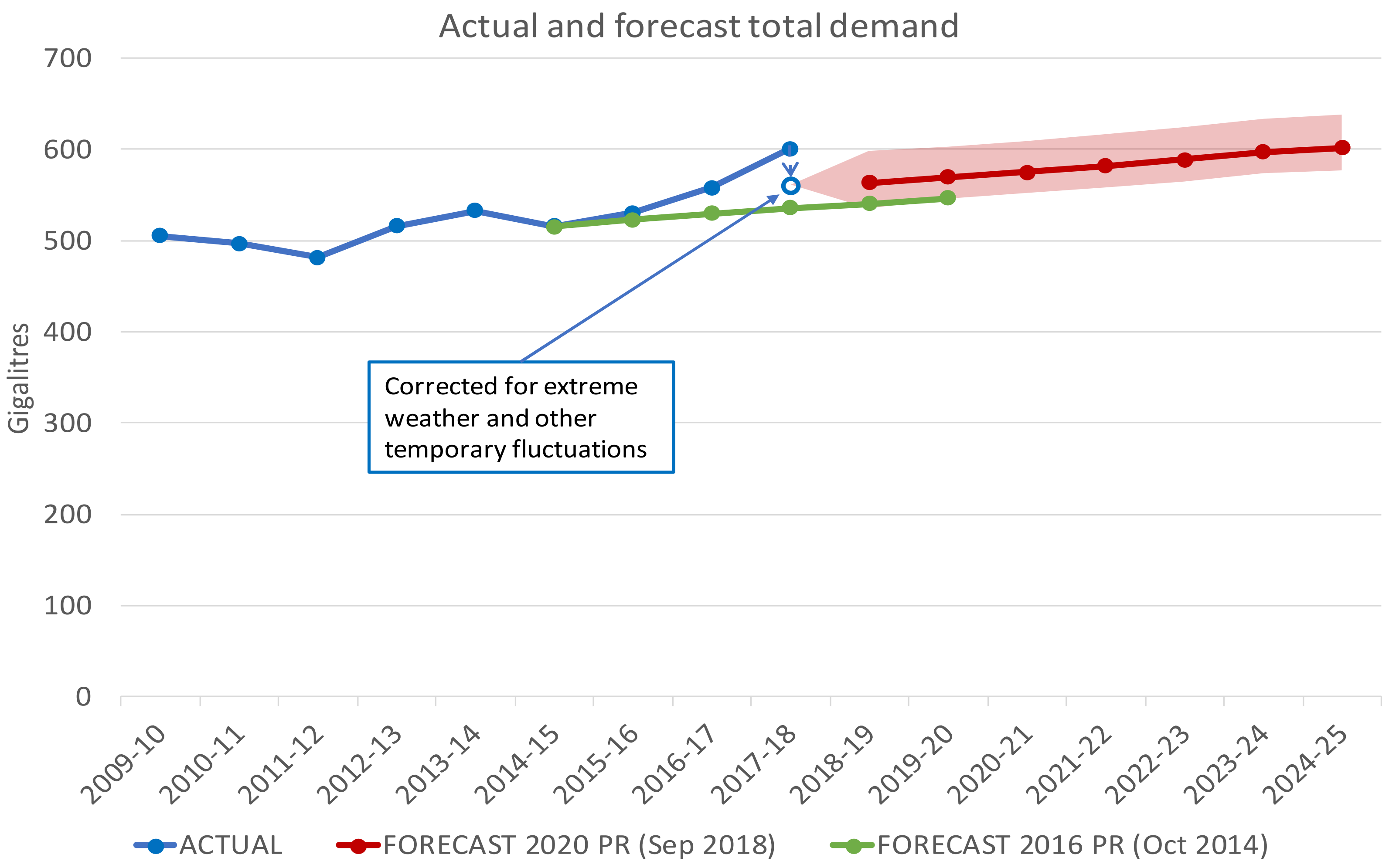


Key drivers of forecast increase:

- Increased dwelling and population projections
- higher than originally estimated impact of the 2016 price decrease
- average weather conditions based on NarClim climate change scenarios instead of long term historical average conditions

Total demand forecast

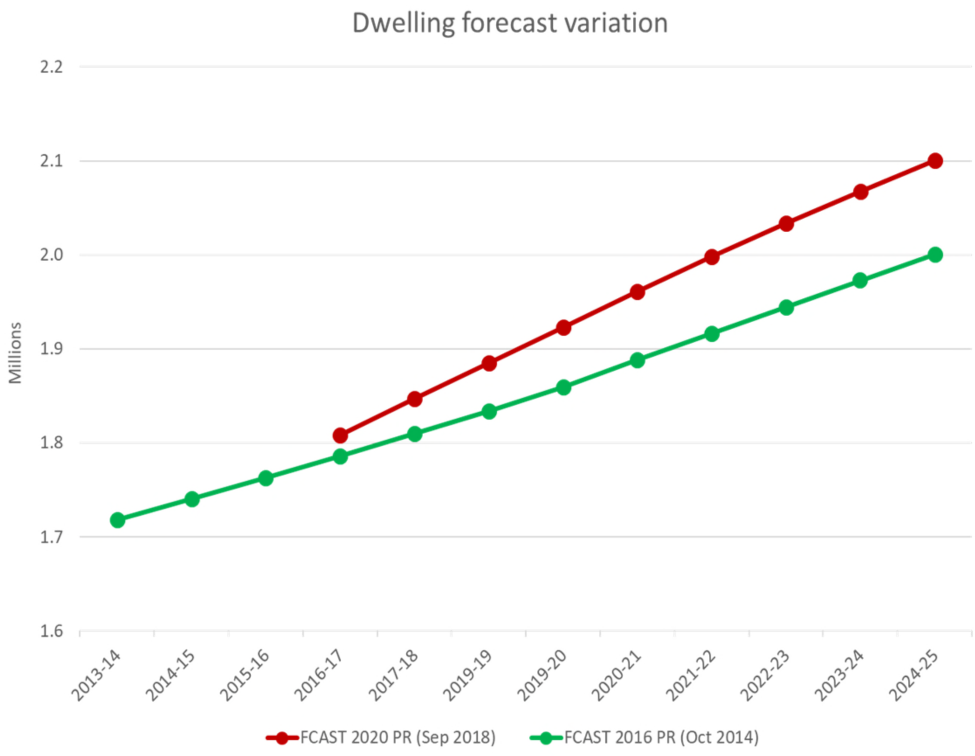
Potential variation due to weather fluctuations



Key assumptions

- Prepared October 2018
- Dwelling growth as per Department of Planning, Industry and Environment projections, adjusted to Sydney Water’s area of operations
- Average weather conditions as per NARCLiM climate change projections for 2020-40
- No change in real water usage price
- Average real losses of 118 ML/day over the determination period
- No water restrictions
- Western Sydney Airport not included

Dwelling forecasts



- 2016 PR: Actuals as at June 2014 and growth forecasts by Department of Planning and Environment released 2012 with minor (upward) adjustments
- 2020 PR: Based on actuals as at June 2018 and growth forecasts by DPE released in 2016

Forecasting approach

by water balance component

					Approach
Water supplied	Revenue water (89.9%)	Billed metered consumption	Res	65.1%	→ econometric models
			Non-res	24.2%	
		Billed unmetered consumption		0.7%	→ apply models for billed metered
	Non-revenue water (10.1%)	Unbilled metered consumption		0.1%	→ historic averages
		Unbilled unmetered consumption		0.6%	→ historic averages
		Unauthorised consumption		0.1%	→ 0.1% of total
		Customer meter underregistration		1.8%	→ 2% of billed metered
		Real losses		7.5%	→ Economic Level of Leakage

(percentages refer to 2016-17)

Residential

Residential

High level approach

- Segment residential properties
- Estimate panel regression model for each segment
- Use regression model to forecast average demand for segment under average weather conditions and assumed price
- Multiply by forecast number of dwellings
- Implemented by delivery system

Residential

History

- Segmentation/panel regression approach first used in a study of the price elasticity of water demand in Sydney in 2010
- Re-estimated and implemented in a forecasting model in 2011 for the 2012 price determination
- Endorsed by expert panel and IPART in 2012 determination
- Models updated in 2014 for 2016 determination
- **Models updated in 2018 for 2020 determination**
- Regression analysis carried out by Associate Professor Vasilis Sarafidis, Dept of Econometrics and Business Statistics, Monash University
- Data preparation and implementation of regression models in a forecasting model carried out by Sydney Water staff
- **Peer reviewed in 2015 and 2019 by [redacted], Sapere Research Group**

Segments

- Segment by
  - Dwelling type
  - Pre/post BASIX
  - Availability of recycled water (RCLD)
  - Owner occupied or tenanted
  - Lot size
  - Number of units
- 34 segments

PROPERTY TYPE	BASIX	RCLD	TENURE	LOT SIZE (m^2)	#UNITS	SEGMENT #
SINGLE DWELLINGS	PRE	NO	OWN-OCC	<=332	NA	1
				333-508	NA	2
				509-662	NA	3
				663-870	NA	4
				871-1262	NA	5
				>1262	NA	6
			TENANT	<=332	NA	7
				333-508	NA	8
				509-662	NA	9
				663-870	NA	10
				871-1262	NA	11
				>1262	NA	12
		YES	OWN-OCC	NA	NA	13
			TENANT	NA	NA	14
	POST	NO	OWN-OCC	NA	NA	15
			TENANT	NA-	NA	16
		YES	OWN-OCC	NA	NA	17
			TENANT	NA	NA	18
VERTICAL STRATA UNITS	PRE	NA	NA	NA	2	19
				NA	>2	20
	POST	NA	NA	NA	2	21
				NA	>2	22
TOWNHOUSE STRATA UNITS	PRE	NA	OWN-OCC	NA	2	23
				NA	>2	24
		NA	TENANT	NA	2	25
				NA	>2	26
	POST	NA	OWN-OCC	NA	2	27
				NA	>2	28
		NA	TENANT	NA	2	29
				NA	>2	30
FLATS	PRE	NA	NA	NA	NA	31
	POST	NA	NA	NA	NA	32
DUAL OCCUPANCIES	PRE	NA	NA	NA	NA	33
	POST	NA	NA	NA	NA	34

Regression model specification

$$\ln c_{it} = \alpha \ln c_{it-1} + \beta_1 \mathrm{I}(\Delta p_{it-1} \geq 0) p_{it-1} + \beta_2 \mathrm{I}(\Delta p_{it-1} < 0) p_{it-1} + \boldsymbol{\beta}' \boldsymbol{x_{it}} + u_{it}$$

$$u_{it} = \eta_i + \varepsilon_{it}, |\alpha| < 1$$

- $\ln c_{it}$: (natural logarithm) of mean daily consumption of property *i* in quarter *t*
- p_{it}: real water usage price in quarter *t*
- $\mathrm{I}(\Delta p_{it-1} \geq 0)$ equals 1 if condition in brackets is met and 0 otherwise

  $\mathrm{I}(\Delta p_{it-1} < 0)$ equals 1 if condition in brackets is met and 0 otherwise

  Therefore, β_1 coefficient measures impact of price increases and β_2 coefficient measures the impact of price decreases, allowing testing for asymmetric price effects.

Regression model specification
Continued

$$\ln c_{it} = \alpha \ln c_{it-1} + \beta_1 \mathrm{I}(\Delta p_{it-1} \geq 0) p_{it-1} + \beta_2 \mathrm{I}(\Delta p_{it-1} < 0) p_{it-1} + \boldsymbol{\beta}' \boldsymbol{x_{it}} + u_{it}$$

$$u_{it} = \eta_i + \varepsilon_{it}, |\alpha| < 1$$

- $\boldsymbol{x_{it}}$ is a vector capturing pseudo dummy variables for season and the following weather variables:
  - Average daily maximum temperature anomaly
  - Average daily rainfall anomaly
  - Average daily (pan) evaporation anomaly
  - Number of days with max temperature > 30C
  - Number of days with rainfall > 2mm
  - Number of days with max temperature > 40C
  - Longest run of days without rainfall
  - Longest run of days with rainfall ≤ 1mm
  - (also tested but not included in final specification: days > 35C; max run ≤ 2 mm; max run > 30C; max run > 35C; max run > 40C)
- u_{it}: error term consisting of a time invariant, individual specific component, η_i, and random noise component, ε_{it}.

Data Sources

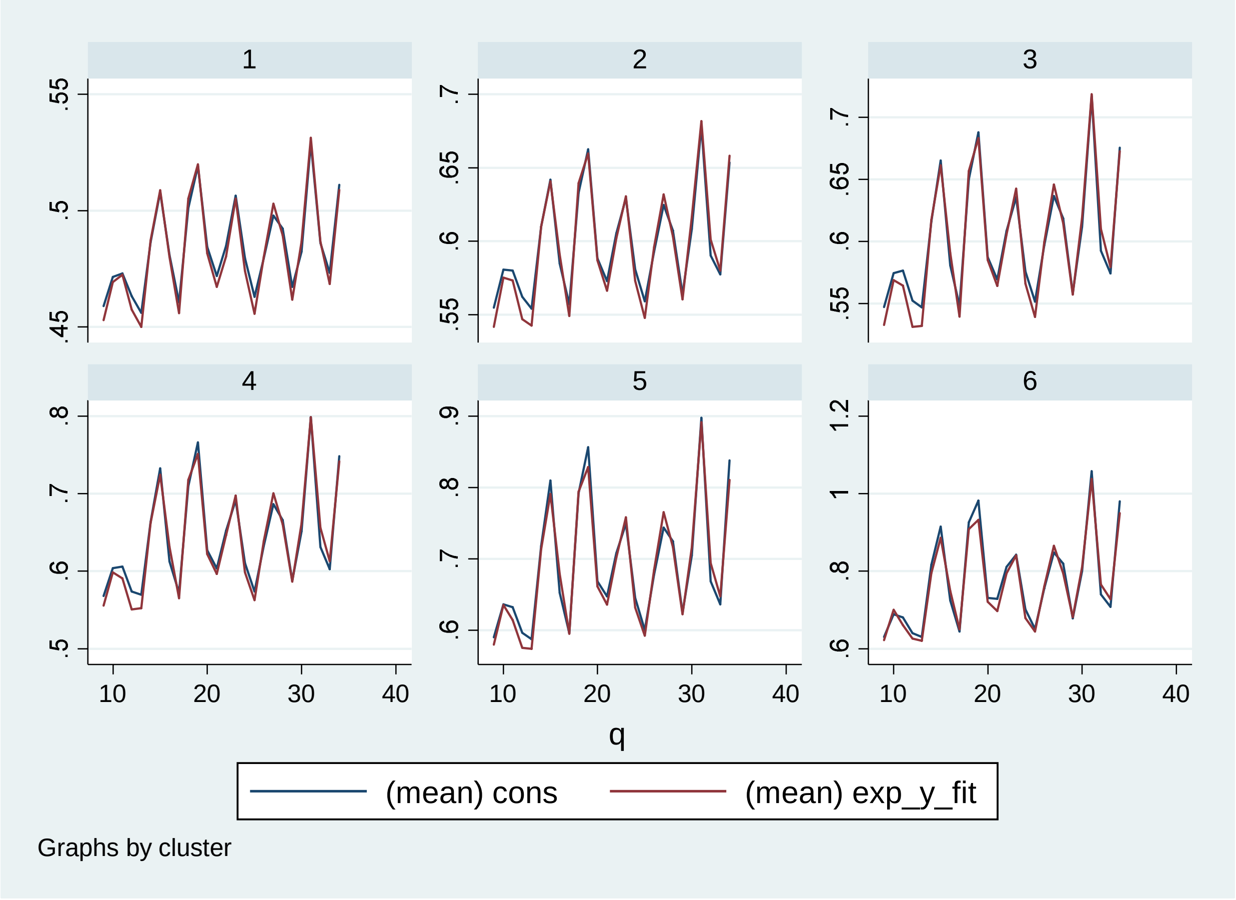
- Quarterly meter reading data, property classification history, lot size, number of units and tenure proxy variable taken from Sydney Water's systems.
- BASIX status approximated using property type classification effective date.
- Gridded data (approx. 5x5 km grid cells) on daily maximum temperature, rainfall and (pan) evaporation data obtained from Bureau of Meteorology
- Each property is matched to a grid cell based on its location.
- For each property and quarterly meter read, daily weather data for the gridcell in which the property is located and the dates covered by that meter read are used to calculate the values of the weather variables. This is to capture the variation in weather conditions between properties due to variation in location and meter reading dates. (Each quarter it takes about 10 weeks to read all meters.)

Estimation

- 26 quarters of data (Q1 2011-12 to Q2 2017-18)
- Ordinary Least Squares not suitable due to endogenous regressors
- Coefficients estimated using a Generalised Method of Moments estimator for dynamic panel models as developed by Arellano, Bond and others.
- Estimation performed using Stata, xtabond2 command.

Results

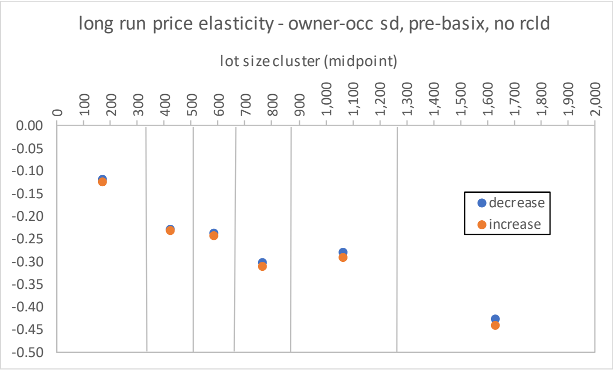
Example: in sample fit, segments 1-6



Results

Price elasticity

- Coefficients for price decreases generally a bit smaller but difference is not statistically significant or material, eg:



- Largest price elasticities for owner occupied single dwellings
- Elasticities tend to be smaller for tenanted and multi-residential dwellings
- Weighted average price elasticities (long run):

	Price decrease	Price increase
Single dwellings	-0.212	-0.218
Multi-dwellings	-0.058	-0.063

Implementation

- To implement the regression models in a forecasting model and generate a forecast requires a number of additional steps. At a high level, these are:
  - Adjust the models to quarterly apportioned consumption:
    - Select every property that is classified as residential as at December 2018 and has at least 5 complete quarters of apportioned consumption data.
    - For each selected property
      - determine the segment it belongs to
      - using the regression model for the segment it belongs to, hindcast the log of apportioned quarterly consumption over the period July 2011 to Dec 2017
      - Use residuals to estimate its property specific constant term (η_i)
  - For each selected property, forecast the log of quarterly apportioned consumption by entering assumed price and long term average weather conditions at its location (see later section for more detail on how average weather conditions were defined)
  - Convert each property's forecast from logs to levels and apply bias correction factor for the conversion
  - Average the forecasts for the individual properties by delivery system, dwelling type, BASIX status and availability of recycled water.
  - Multiply these by the forecast number of dwellings by delivery system, dwelling type, BASIX status and availability of recycled water

Non-residential

Non-residential

High level approach

- Segment properties by
  - Consumption (6 highest users as at 2011)
  - Water efficiency program participation
  - Segment remaining properties by property type:
    - Industrial
    - Commercial
    - Government & Institutional
    - Agricultural
    - Commercial & Industrial strata units
    - Standpipes
    - Other
- Other refers to a number of property types which do not obviously fit in to residential or non-residential, eg “occupied land”. Their consumption is relatively small (<1% of total demand) and constant over time.
- Develop time series of average demand by each segment in each system using monthly apportioned consumption data.
- Deseasonalise average demand and fit a time series regression model to the deseasonalised demand modelling the remaining variation as a function of weather anomalies and trend.
- Use the regression model to estimate a weather corrected deseasonalised average demand for each segment as at 2011-12.
- To forecast average demand for each segment, assume weather corrected deseasonalised average demand remains constant over the forecast period.
- Seasonalise the forecast and then multiply by the forecast number of properties in each segment.

History

- Approach first used in 2011 for the 2012 price determination. Not differentiated by delivery systems at that time.
- 2013: Introduced separate models for each delivery system.
- 2016: Price determination: Added a price elasticity estimate.
- 2018: Addition of a densification factor, price asymmetry factor removed and corrections to “Other”.

Model specification

$$c_{ijt} = \left[\bar{c}_{ij} + f(T_t, R_t, E_t)\right] \times s_{ijt} \times (e \times \dot{p}_t) \times d_{jt} \times n_{ijt}$$

$$f(T_t, R_t, E_t) = \beta_{1,ij} \times \Delta T_t + \beta_{2,ij} \times \Delta R_t + \beta_{3,ij} \times \Delta E_t$$

- This is the general specification as used for all segments except Top 6 and Other.
- c_{ijt} is the forecast consumption by segment *i* in system *j* in month *t*
- $\bar{c}_{ij}$ is the deseasonalised average consumption of segment *i* in system *j* under average weather conditions. It was estimated using the time series regression model.
- $f(T_t, R_t, E_t)$ is a "weather correction" which is a function of the maximum temperature anomaly (ΔT_t), rainfall anomaly (ΔR_t) and evaporation anomaly (ΔE_t) at time *t*. Coefficients are taken from the regression model. Anomalies are calculated relative to the long term (30 year) average.
- This weather correction was used to incorporate climate change in the forecast. The correction is based on the difference between the forecast average conditions with climate change and the 30 year average which is the base for the anomalies. See below for more detail.

Model specification

Continued

$$c_{ijt} = \left[\bar{c}_{ij} + f(T_t, R_t, E_t)\right] \times s_{ijt} \times (e \times \dot{p}_t) \times n_{ijt} \times d_{jt}$$

$$f(T_t, R_t, E_t) = \beta_{1,ij} \times \Delta T_t + \beta_{2,ij} \times \Delta R_t + \beta_{3,ij} \times \Delta E_t$$

- s_{ijt} is the multiplicative seasonal factor for segment *i* in system *j* for month *t*, estimated using seasonal decomposition (ratio-to-moving average method).
- e is the non-residential price elasticity (-0.264) and $\dot{p}_t$is the relative change in the real water usage price at *t* (base year: 2015-16).
- n_{ijt} is the forecast number of properties in segment *i* in system *j* for month *t*.
- d_{jt} is a densification factor which is the ratio of the most recent population forecast and the forecast population as available when the models were estimated in 2013. It was introduced to capture that while population growth (a proxy for workforce growth) has accelerated since the models were estimated, non-residential property growth has not.

Model specification

Continued

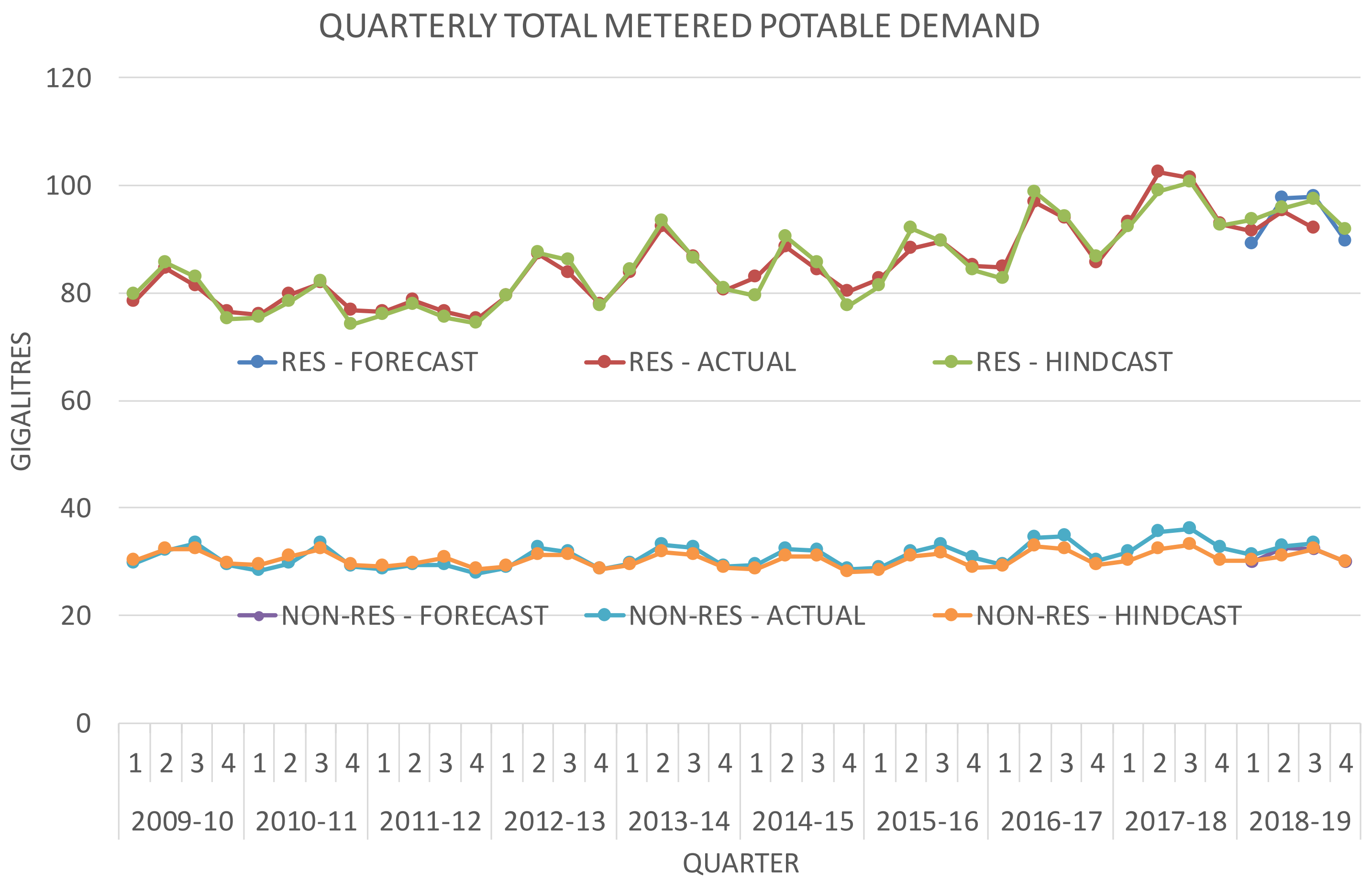
$$c_t = \sum\nolimits_{ij} c_{ijt} + top6_t + other_t + bias\ correction$$

- To calculate the total demand forecast for month *t*, we sum the forecast for all segments in all systems ($\sum_{ij} c_{ijt}$) and add the forecast demand by Top 6 properties ($top6_t$) and Other properties ($other_t$) and a $bias\ correction$.
- Top 6 demand is forecast using a specific model for each Top 6 user based on historical consumption and known water saving initiatives, eg introduction of the Rosehill-Camelia recycled water project.
- Consumption by other is relatively small and constant and is forecast to be constant at 4 GL/year based on historical data.
- When we applied the updated model in hindcast mode, it underestimated observed demand by about 1.3 GL/year, on average. This average underestimate is added to the forecast and referred to as $bias\ correction$ in the model specification above.

Hindcast performance

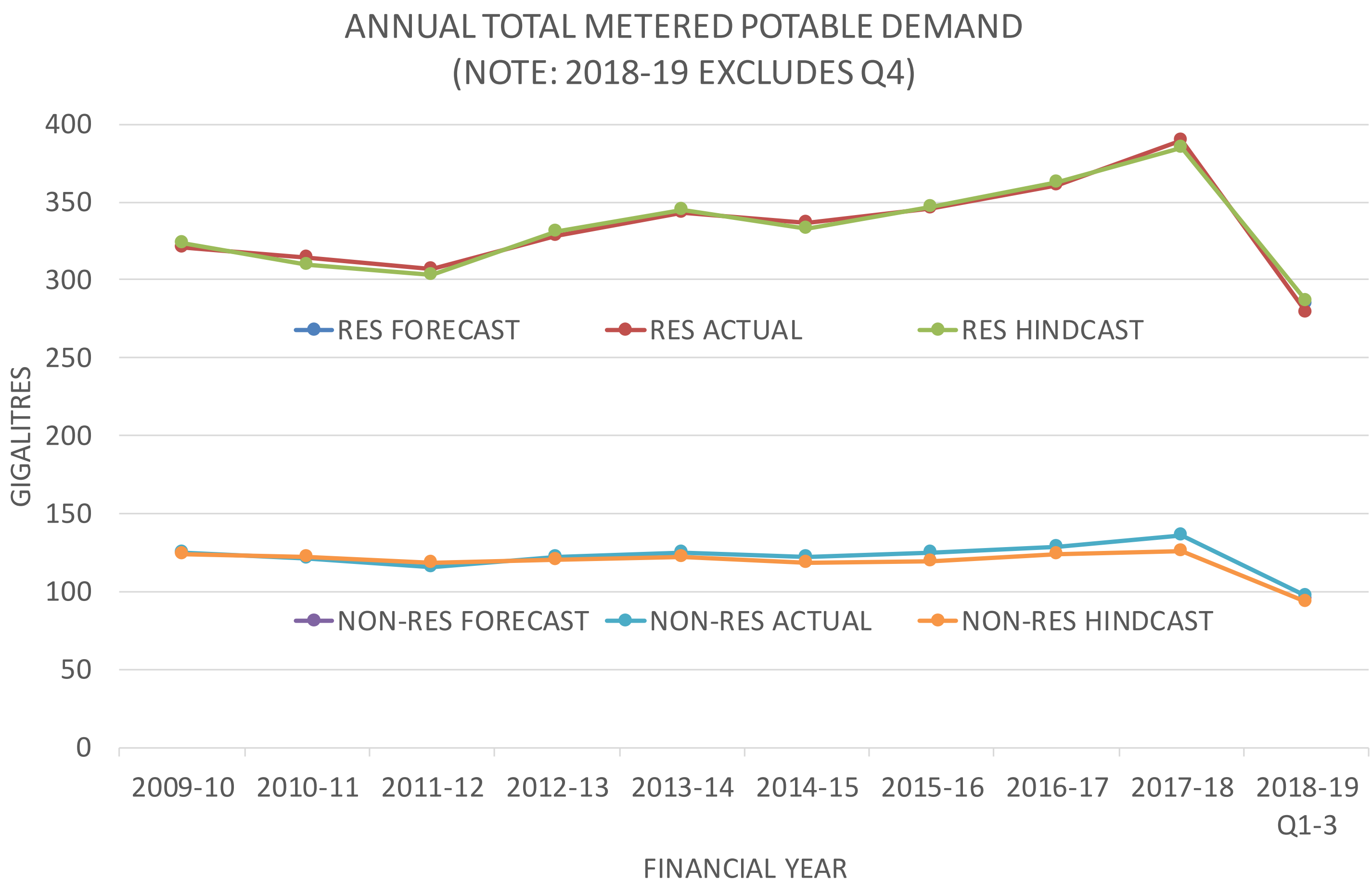
Performance

Hindcast vs actual



Performance

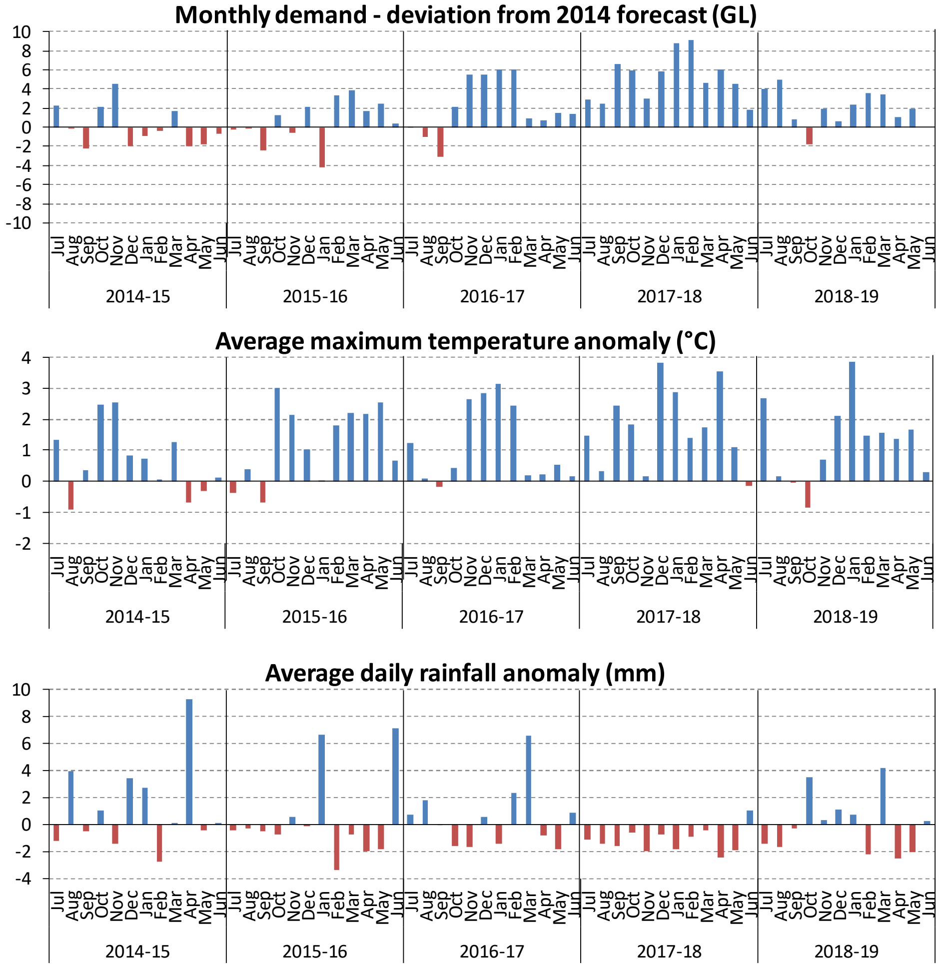
Hindcast vs actual



Incorporating climate change

Demand and weather

- Weather cannot be predicted over the timeframe required for the price review (5-6 years from time of preparation to end of price path)
- Forecast therefore assumes average weather conditions
- Deviations from forecast correlate strongly with deviations from average weather conditions



Weather variables based on weighted average data from BoM stations at Sydney Airport and Prospect Reservoir (or nearby Horsley Park). Anomalies calculated relative to 30-year average to June 2010.

Defining average weather

- Long term average weather conditions were previously defined as the 30 year average values, a standard averaging period in climatology.
- However, this approach may not be valid in the presence of climate change. Climate change means weather variables are not stationary but trending. For example, climate change causes temperatures to fluctuate around an upward trend, not around a stationary average.
- To address this issue we have defined average weather conditions using the regional climate projections as developed by the NARCLiM project.

NARCLiM

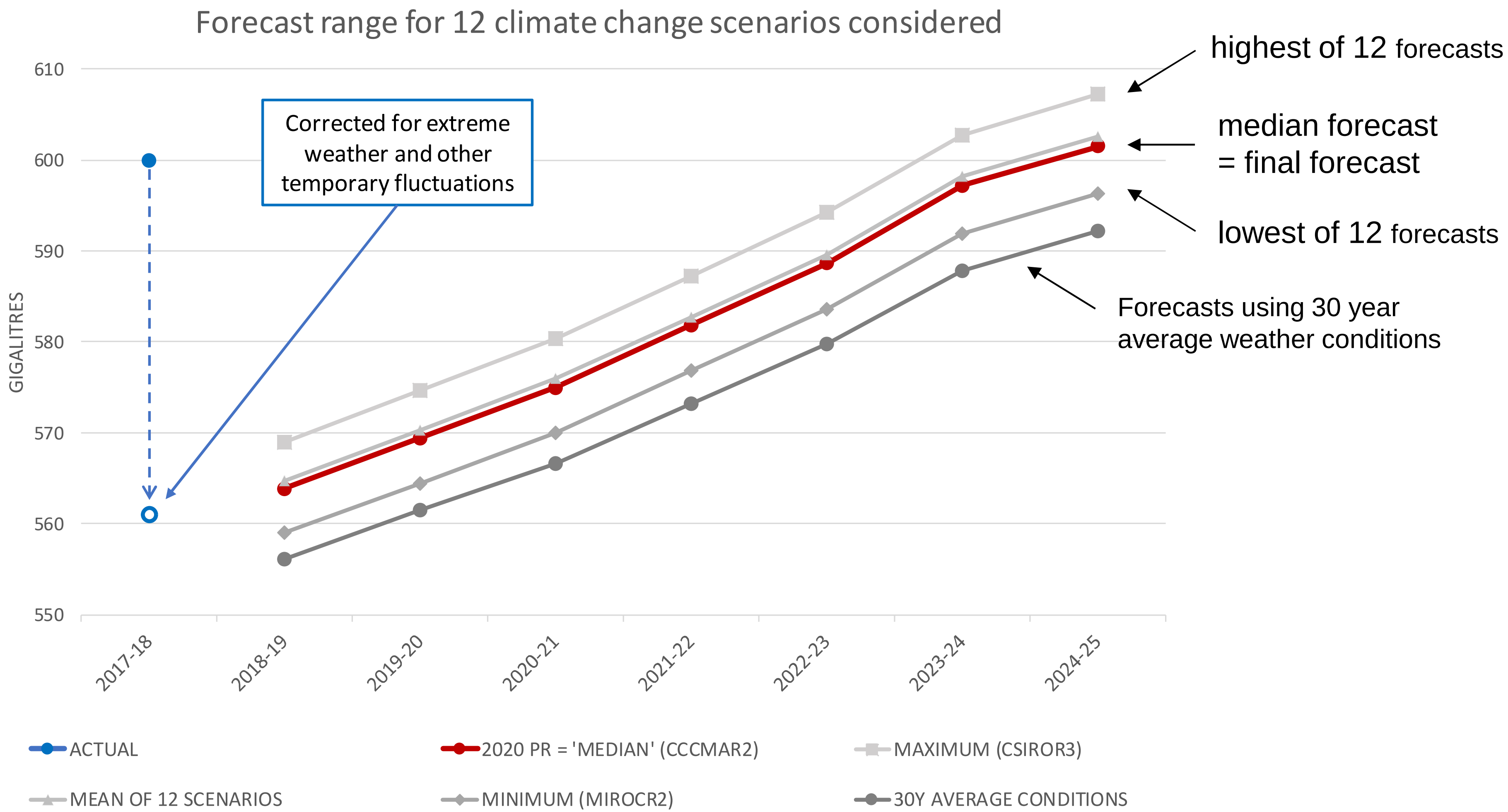
Downscaling global climate model results

- NARCLiM is research partnership between NSW Office of Environment and Heritage and Climate Change Research Centre at University of NSW (UNSW)
- NARCLiM aims to provide planners with high resolution projections of the impacts of climate change
- NARCLiM takes the results of global climate models which produce averaged results over large areas and uses dynamical downscaling methods to translate this into projections for smaller areas of approximately 10x10km
- See https://climatechange.environment.nsw.gov.au/Climate-projections-for-NSW/About-NARCliM/
- UNSW interpolated the NARCLiM outputs to a 5x5 km grid for Sydney Water. This is consistent with the grid as used for the historical BoM data which was used to estimate the regression models. UNSW also added estimates of pan evaporation which is currently not a NARCLiM output using a probabilistic model conditioned on the maximum temperature, rainfall and gridcell.

Applying the NARCLiM projections

- NARCLiM used results from 4 global models as input and used 3 different approaches to downscaling to produce a total of 12 projections.
- We used the projections for the 2020-2040 period to develop our demand forecast.
- 12 demand forecasts were produced, one for each of the 12 NARCLiM projections. The median of these 12 forecast was taken as our final forecast.
- This forecast is about 1.5% higher than the forecast that results from using the 30 year average to define average weather – see next slide.

Forecasts under climate change



Billed unmetered and non-revenue water

Billed unmetered and non-revenue

- Billed unmetered consumption: estimated by applying the models for billed metered consumption to unmetered properties
- Unbilled metered: assumed constant at 320 ML/year based on historical averages
- Unbilled unmetered: assumed constant at 3,500 ML/year based on historical averages
- Top up of recycled water systems: forecast on basis of forecast number of properties with recycled water and historical average top up rates from Rouse Hill scheme
- Unauthorised: forecast as 0.1% of forecast total demand
- Underregistration: forecast as 2% of forecast billed metered demand

Billed metered and non-revenue

Real losses (leakage)