Australasian Groundwater & Environmental Consultants

1 September 2022

South32 Ltd
Illawarra Metallurgical Coal
Port Kembla Road
Port Kembla 2505
New South Wales Australia

Attention: Cody Brady
via email: Cody.Brady@south32.net

Dear Cody,

Appin Mine Groundwater Model Peer Review

1 Introduction

The Appin Mine is an existing underground coal mine located approximately 25 kilometres north-west of Wollongong. Appin Mine is owned and operated by Illawarra Metallurgical Coal (IMC), a subsidiary of South32. The Appin mining operations in the Bulli Seam are known as the Bulli Seam Operations (BSO) and undertaken in accordance with Project Approval 08_0150. IMC is currently extracting Longwall 709 in Area 7 and Longwall 905 in Area 9 and has received Extraction Plan (EP) approval for Longwalls 709, 710A, 710B, 711 and 905, which are referred to as 'the Project' in reporting.

Heritage Computing (2009) developed the first groundwater model for Appin Mine. The model was updated in 2020 by SLR Consulting Australia Pty Ltd (SLR) to simulate groundwater impacts for the Project. SLR have recently revised the 2020 numerical model following a review from the then NSW Department of Planning, Industry and Environment (DPIE) (2021) Biodiversity and Conservation Division (now NSW Department of Planning and Environment).

This letter report provides a peer review of the updated 2022 groundwater modelling undertaken by SLR for the EP. Australasian Groundwater and Environmental Consultants Pty Ltd. (AGE) undertook this review at the request of South32 Limited (South32).

2 Methodology

The objective of the peer review was to assess the conceptual and numerical models described in the groundwater assessment report against available guidelines for groundwater modelling.

The following report was supplied for the review:

- Appin Mine Extraction Plan. Groundwater Impact Assessment. SLR. V8.0. August 2022.

The reviewer also attended three videoconference meetings with representatives of SLR and South32 to discuss the model calibration (14/06/2022), and model predictions (7/06/2022, 27/05/2022).

The other documents used during this peer review were:

- Barnett, B, Townley, LR, Post, V, Evans, RE, Hunt, RJ, Peeters, L Richardson, S, Werner, AD, Knapton, A, & Boronkay, A (2012), *Australian groundwater modelling guidelines*. Waterlines report, National Water Commission, Canberra (herein referred to as the AGMG).
- Commonwealth of Australia (CoA), (2018), Information guidelines for proponents preparing coal seam gas and large coal mining development proposals, Commonwealth of Australia, May 2018.
- Middlemis H and Peeters LJM (2018), Uncertainty Analysis – Guidance for groundwater modelling within a risk management framework. A report prepared for the Independent Expert Scientific Committee on Coal Seam Gas and Large Coal Mining Development through the Department of the Environment and Energy, Commonwealth of Australia 2018.
- Murray–Darling Basin Commission (MDBC) 2001, Groundwater flow modelling guideline, report prepared by Aquaterra, January 2001.
- Doherty, J. and Moore C., (2021). Decision Support Modelling Viewed through the Lens of Model Complexity. A GMDSI Monograph. National Centre for Groundwater Research and Training, Flinders University, South Australia.

3 Review and discussion

The following sections review the SLR report against the Australian groundwater modelling guidelines (AGMG) as well as the other documents noted above.

3.1 Objectives

The objectives of the groundwater assessment were to describe the existing hydrogeological environment, and assess the potential impacts of mining on the groundwater regime. The scope of works developed by SLR was designed to meet this objective and respond to comments from DPIE including:

- providing background information on the site setting and conceptual groundwater model;
- calibrating the numerical flow model suitable to predict Project impacts in accordance AGMG and MDBC (2001), including improving previous mismatches between modelled and observed groundwater levels;
- predicting cumulative impacts on the groundwater regime from the Project and surrounding activities;
- calculating baseflow/leakage impacts, drawdown, groundwater interception and incidental water impact; and
- providing recommendations for ongoing monitoring and establishing groundwater triggers.

The numerical model has been employed for a wide range of purposes, which is typical for most models utilised to assess the impact of mining in NSW. This is driven by the requirements of the NSW Aquifer Interference Policy (AIP) that requires a wide range of potential impacts to be estimated. The key impacts to be predicted are mostly differences between two models rather than absolute values. This is preferable as predictive differences may be less uncertain than absolute values (Doherty and Moore 2021). One of the main purposes of the model is to estimate the potential impact on private water supply bores. This objective is not directly stated within introductory sections of the report, but relevant information is provided in latter sections that outline modelling predictions on this key receptor.

3.2 Conceptual model

The groundwater assessment report contains some 30 pages of text and graphics describing the available hydrogeological datasets, and four pages summarising the conceptual model of the groundwater regime. The Hawkesbury Sandstone is identified as the main aquifer with modest yields and relatively good water quality. Quaternary alluvium also forms a thin and sporadic aquifer where this occurs along some creek lines.

Upon review of the report the Project area and surrounds appears to be a challenging area to gather hydrogeological data. This is likely due to a range of factors including the significant depth of the Project, and restrictions to land access due to government and private land ownership. Despite these challenges a good monitoring network appears to have been built up over time. The groundwater monitoring network primarily utilises vibrating wire pressure sensors (VWPs) that are sealed within boreholes at different depths to measure pore pressure over time within the key hydrostratigraphic units.

This is considered a logical and appropriate monitoring methodology to measure changes within the groundwater regime over time given the relatively deep geological setting. The monitoring points are located across areas of historical mining, as well as within or adjacent to the Project area.

The length of the groundwater monitoring record varies depending on location, with some sites having a data record of more than 10 years. This length of the baseline monitoring is good, as it exceeds the eight-year Project life in some areas allowing the model calibration period to exceed the length of the future prediction period.

The climate during the monitoring period has included years of typically average rainfall, and a short but intense drought period between 2016 and 2020. This climate variability is reflected in measured groundwater levels at many monitoring sites. Declining groundwater levels attributed to mining induced depressurisation have also been recorded at some monitoring sites within the monitoring network, with the impact of mining reducing vertically above the longwall panels. The varying climate conditions and the recorded mining impacts vertically through the strata over the monitoring period provides an information-rich dataset for history matching as part of the numerical modelling.

Observed groundwater inflows to the active mining areas at Appin are not provided within the SLR report. It is acknowledged groundwater inflows occurring in underground mines can be challenging to estimate as water is pumped into the operations from the surface for dust suppression and machinery use. However, a simple water balance can often identify likely ranges of groundwater inflows that can be used to inform numerical modelling history matching.

The geology and associated hydrostratigraphic units occurring within the Quaternary, Triassic and Permian formations are described within the report, supported by a geological map and a vertical geological section. The groundwater assessment report does not contain any information on the measured hydraulic properties of the key hydrostratigraphic units. Whilst it is acknowledged the site setting would make in-situ measurement of hydraulic properties challenging, and properties can be inferred based on experience, the report would be improved by including a summary of any available hydraulic test results.

There is also no site-specific data available on the influence of longwall mining on the hydraulic properties of strata overlying Appin Mine. SLR note hydraulic conductivity is increased by 2 to 3 orders of magnitude within the goaf and fractured zone based on data from surrounding mines. In the absence of any measurements at Appin, SLR utilise common empirical methods to estimate the height of the fractured strata above the longwall mining areas.

Potentially sensitive receptors reliant on the groundwater regime within the Project area are groundwater fed creeks and rivers, and private water supply bores. A good summary of the location and details of registered water supply bores is provided within the SLR report, based on information within the NSW government groundwater database. Where the information in the government database is lacking, SLR identify the formation the bores are drawing water from based on the recorded bore depth. It is not stated within the groundwater assessment report if a survey of private properties within the Project area has ever been conducted to confirm the location and details of water supply bores. This would be an appropriate future step within the likely impact zone to identify the exact location of registered and any unregistered bore in use.

The report provides information on the main creek and rivers including flow gauging plots and a summary of water quality. There is no discussion on the nature of any groundwater-surface water interactions provided within the report. Whilst there is no discussion on this topic, the water level contour map included for the Hawkesbury sandstone does provide an indication of creek and river reaches that could be either losing, and/or gaining groundwater. It is not clear how this map was generated, but if numerical modelling was utilised this should be acknowledged. Simple comparisons between measured groundwater levels and gauged river levels would also assist with this interpretation.

There is also no comment on the potential for groundwater dependent ecosystems such as riparian vegetation to occur in the Project area. This is a common component of groundwater assessment reports so it is unclear why this information is omitted.

The report notes that after closure of the mine groundwater levels will recover, and that the pH could reduce over time, resulting in increases in the concentrations of metals such as zinc, iron and nickel. This statement in the report is not supported by site specific data, and the scope of work did not include an assessment of post mining impacts, so it's unclear why this speculation was included in the document.

Cumulative impacts on the groundwater regime in the region are significant, with the presence of other aquifer interference activities (operating or closed underground coal mines and coal seam gas extraction) as well as a network of private water supply bores that are constructed within strata that overlie and surround the Project. These are all described as much as possible within the groundwater assessment report utilising public domain information.

3.3 Numerical model setup

SLR constructed a large regional numerical model to represent the conceptual model of the area. The model was constructed using the MODFLOW-USG software, an industry standard package for this application. The model utilises a Voronoi shaped cells, which are refined to 100 m in areas of interest such as creek and longwall panels, with 50 m cells used to represent shafts. The model has 18 separate layers representing the main hydrostratigraphic units occurring in the region. Key aspects of the conceptual model are represented in the numerical model including rainfall recharge, evapotranspiration, stream stage height, private water bore pumping and coal seam gas pumping with standard MODFLOW packages and approaches.

Rainfall recharge is represented with the RCH package that applies a fixed percentage of the total annual rainfall as recharge to the water table. This is a commonly adopted approach in regional models, but one which introduces an averaging effect which means the model cannot closely replicate water level variability due to shorter duration rainfall events. Pumping from water bores and coal seam gas bores is represented using the WEL package with assumptions on pumping rate based on information in the public domain or plausible assumptions.

Longwall mining is represented with advancing DRN cells, with a drain conductance of 100m^2/day. The TVM is used to represent fracturing above the longwall panels with vertical and horizontal hydraulic conductivity increases. The specific yield within the mined coal seam was changed to 10% to represent formation of the goaf, which is a plausible assumption of the residual air volume within the seam post mining.

Aspects of underground mining are also represented including longwall mining, strata fracturing and underground water storage. The total cell count is about individual 890,000 cells, and quarterly stress periods resulting in a model run time understood to be about one hour. The model is therefore a relatively large and complex model, but with a modest simulation time that allows history matching and uncertainty analysis.

Changes were also made to the model to address comments from DPE regarding the height of fracturing and the calibration to head measurements. To address these comments the height of fracturing was updated and a surface fracturing zone represented in the model (this is discussed further in Section 3.4). The period of groundwater levels utilised in the model was also extended to include data from 2010 to 2021.

3.4 Calibration

The 2022 model was calibrated using industry standard optimisation software (PEST++) through adjustable parameter zones. Regularisation was not used.

The model was setup with uniform hydraulic property values applied in zones where each hydrostratigraphic unit occurs. Model layers 1 and 2 contained zones representing the outcropping formations, with layers 3 to 18 representing single hydrostratigraphic units with uniform hydraulic properties. The actual hydraulic properties will be more heterogenous than represented in the model layers. The uniform parameter values in the model imposes a lack of flexibility during calibration that means the model cannot replicate every nuance in heads and drawdown measurements that are driven by localised hydraulic properties. The model is however able to replicate an averaged fit to the measurement data, meaning the average head trends can be replicated. Good replication of water level trends on a bore-by-bore level is not possible with the model setup which means a larger misfit must be accepted. This is not a deficiency of the model per se; it is simply something that needs to be considered upon review.

However, if the model was used as a predictive tool using linear analysis using this parameter scheme, the posterior uncertainty predicted would be overly constrained and under-representative of potential impacts. The future use of many pilot point multipliers to allow model parameters to vary more spatially during the history matching process would remedy this.

Whilst there are no measurements of hydraulic properties provided in the report, the calibrated parameter ranges appear plausible based on experience in similar geological settings.

The range of parameters used during calibration, and the sensitivity of those parameters are not presented by SLR. It is unknown if any of these parameters have hit their bounds or not, the latter implying some structural defect is not represented in the model, or an inappropriate range constrained by PEST. It is noted that the hydraulic conductivity of the layers above the Bulli Seam are at the low end of the range of anticipated values; in particular the Stanwell Park Claystone which is very low and almost one order of magnitude lower than any other layer at $7.3E^{-7}$m/day.

The report states Kh and Kv above the longwall were adjusted during the calibration process. However, it is understood they were adjusted using a fixed factor which was not specifically explored using PEST++. As a result, there is no information available on the sensitivity of the calibration/predictions to these important parameters. The report states Kh and Kv in lower zone of connected fracturing was increased according to the methodology to determine permeability in the fracture zone provided by Guo (2007). The Guo (2007) equation suggests Kh increases of 15-1000x the host value in the first 100m above the longwall, and Kv increases 2-40x. Increases to Kh and Kv from the zone 100-200m caused the models to decalibrate (failure to match head and inflows) meaning the magnitude of permeability changes recommended by Guo were not represented in the SLR model. The observation data supports the lack of vertical connectivity through the fracture zone as groundwater levels become less obviously impacted by mining as height above the longwall mining increases. There are two possibilities to explain this outcome, firstly, the model is correct and fracturing is very limited within the zone from 100-300m above the longwall, or secondly fracturing is more extensive, but the interconnection of fracture networks is poor, and regional throughflow buffers drawdown.

Calibrated specific storage (Ss) values appear reasonable; the geometric mean is approximately 1E-06 m^{-1}. Ss is generally constrained to theoretical bounds presented in Rau (2018), while some values in the model extend to the lower bound of about 1E-07 m^{-1}. It is not stated if Ss moved far from the initial values during history matching. This information would help identify whether hydraulic conductivity is more estimable than Ss with the given measurement dataset. It should be noted that according to poroelastic theory Ss is a function of bulk modulus, porosity, and Poissons ratio (Pells, 2017). Considering the properties of sandstone, interburden and coal measures in the region, an Ss value of around 2 E-06 m^{-1} is more likely. This could be validated with triaxial testing data. A higher value of Ss should be considered for sensitivity and uncertainty exercises.

Calibrated recharge appear plausible. Adopted recharge rates are significantly lower than the 2009 study, yet closer to estimates of recharge at surrounding projects. As highlighted, because recharge is applied as a percentage of rainfall it is difficult for the model to reproduce groundwater level highs (prolonged rainfall events that exceed soil storage) and the lows (prolonged rainfall events that are less than soil storage). This means stress periods with zero recharge where accumulating rainfall is consistently less than the soil store and evaporation are not represented in the model. Applying a zero recharge rate during dry periods in the model could temporarily increase impacts. This could be considered in sensitivity scenarios at least. Regardless, most of the measurement data used during the calibration process indicates deep groundwater levels around the Project area, suggesting a slow response to surface water stresses.

The scatterplots, RMS and SRMS statistics appear reasonable. The RMS and SRMS have reduced by approximately half compared to the 2009 modelling which represents an improvement. This is partly due to the extended range in the measurement set in the Bulli seam. There is no discussion on any observations that were removed from the measurement set during history matching. Some discussion on the rejection process would improve the understanding of the history matching process.

As discussed above, the setup of the model layers and uniform zones of hydraulic properties mean the model is unlikely to be able to reach a better statistical fit than was achieved. It is also likely that one of the most significant influences on groundwater levels and drawdown in the model is the hydraulic conductivity of the fracture zone above the historical longwall mining.

Although the hydraulic properties of the fractured zone were not adjusted during the calibration process, it is likely they are highly estimable (they lie squarely in the solution space) due to the available groundwater monitoring data influenced by mining. A better fit to available data might have been achieved if hydraulic properties of the fracture zone were adjustable in each layer and cell during the history matching process. The modelled groundwater levels generally replicate the vertical downward hydraulic gradient observed in the multilevel VWP sensors. Reallocating bores with uncertain screen intervals has also improved the fit with the observation data.

Spatial residuals in the pertinent aquifers/aquitards appear to be between about 10 m to 40 m around the Project area (it is difficult to interpret as the scale on the map is different to the legend). These residuals are primarily due to the models inability to simulate complicated dewatering mine processes and the lack of heterogeneity represented in the hydraulic properties. Vertical gradients are presented and discussed. These are generally well replicated by the model. Further calibration efforts could history match to these differences separately as well as absolute values of heads. This makes vertical hydraulic conductivity more estimable.

Groundwater inflows to the mine workings were not used for calibration, but to verify the predicted inflow. Whilst these values were not included in the report, it is understood the modelled values are close to measured data, which suggests the relatively low increases to hydraulic conductivity above the longwall panels in the model is justified.

Overall mass balance appears reasonable. It is not stated in the report if there is consistency between the steady state and transient model, and if there are any timesteps where percent error is greater than 2%.

3.5 Predictions

Longwall mining is represented using a permeability enhancing multiplier based on surrounding projects and verified using depressurisation signatures from site vibrating wire piezometers and inflow estimates (not presented). Drain cells with a nominally high conductance are used in the Bulli Seam and the TVM package is changed to the estimated height of fracturing above the panels. Representing enhanced fracturing using this approach can be problematic. If there are aquitards with very low vertical hydraulic conductivity close to the longwall, then the resultant enhanced permeability due to the applied factor remains low. These layers then act as a buffer to groundwater depressurisation propagating through the strata vertically. For this reason the use of 'stacked drains' has sometimes been adopted as an alternative approach. Notionally, the calibrated version of the Appin groundwater model can replicate some water level trends that suggest the fracture zone is not well interconnected hydraulically. It should be noted that enhancement of hydraulic conductivity in the fracture zone is much more significant than that represented by the 2009 groundwater model (Heritage Computing, 2009).

Groundwater drawdown at neighbouring landholder bores is presented and discussed. The predictions indicate that water levels at five private water supply bores will be impacted and will require make good agreements.

Forecasts of changes to baseflow is briefly discussed and reported to be "negligible". As discussed, it is unknown if this result is caused by the model under-representing measured baseflow due to low groundwater levels or low river conductance (calibrated values of river conductance are not presented). Fortuitously the model predicts minimal impact at the surface, meaning it is likely the negligible forecasts on baseflow impacts are justifiable.

The predictions are not compared to the Minimal Impact Considerations outlined in the NSW AIP. It is unclear why the predictions are not compared to the AIP thresholds, as this is standard practice for groundwater assessments conducted for mining projects in NSW.

3.6 Sensitivity/Uncertainty analysis

The report presents the results from a model that uses the Tammetta method to calculate the height of the A zone. Because of the configuration and depth of the seven and nine series longwall panels, the resultant height of fracturing is lower than the Ditton a95 surface, meaning the predicted impacts are less.

There is no further discussion of the sensitivity of hydraulic parameters to groundwater impacts. It is important to understand that this model represents one realisation in an infinite number of realisations that can calibrate the model. Different combinations of parameters can produce the same level of fit, but with different predicted impacts. It is best practice to explore sensitivity scenarios that consider the structural defects in the model and try to overcome them to quantify the likelihood of worse case impacts. In particular the sensitivity of the predicted groundwater impacts to the adopted fracturing multipliers should be quantified. Although not published, SLR developed a version of the model with much higher fracture multipliers in the A and B zone. This version of the model produced unrealistic historic depressurisation and inflows. On this basis, assuming recharge and storage is ‘correct’, the modeller could reject this as a worst case outcome.

On top of this, combinations of parameters should be considered (i.e., high fracturing, lower storage, and low recharge) to reject or prove the likelihood that the AIP thresholds will be exceeded (e.g., landholder drawdown, or significant baseflow reduction). When conducting this sort of sensitivity analysis, it is important that the realisations can fit historic water level and inflow measurements. It is possible that realisations can both fit historic measurements, but not predict significant future impacts that exceed AIP thresholds.

4 Australian modelling guidelines

The AGMG outlines a process for evaluating numerical models to determine if they are ‘fit for purpose’. The aim of the guidelines is to provide a more appropriate and consistent approach to model development across the industry. The guidelines include a series of checklists that are used to evaluate and classify models according to their complexity. Attached are a series of four tables that address the check lists within the AGMG, as well as a check list included in the predecessor to the AGMG (MDBC 2001).

Table 1 summaries how well the groundwater assessment complies with recommendations of the AGMG. It is concluded that the model and accompanying report were produced to a high standard, and the outcomes are fit for the intended purpose.

Table 2 summarises the three classes of numerical models outlined within the AGMG. The table shows that the model has elements of a class 2 (impact assessment) and class 3 model (complex simulator). This is a typical outcome for most models used for this type of application.

Table 3 is an additional check list that assesses the model against each of the eight stages recommended by the AGMG. The table shows that the numerical model meets many of the requirements of the AGMG, with the exception of uncertainty analysis which is limited.

Table 4 is a checklist from the predecessor to the AGMG (MDBC 2001) which is used to provide a rating for a model based on how well it implements the recommendations of the guideline. The numerical model commonly achieves an ‘adequate’ or ‘very good’ score against this guidance.

5 Conclusions

The modelling described by SLR has been conducted with a methodology and care consistent with industry standard practice. The 2022 study has improved upon the 2020 study by: including: more deep groundwater level measurements in the history matching process, representing the enhanced permeability above the longwall more appropriately, and exploring an alternate fracture height configuration. The calibrated version is fit for purpose of predicting drawdown due to the proposed mining in the Appin Areas 7 and 9 series panels. The ability of the numerical model to predict impacts on surface water and shallow systems is lower due to underprediction of shallow groundwater levels. Due to absence of parameter sensitivity or predictive uncertainty analysis it is unknown if there are other permutations of the model that replicate measurement data but cause more substantial impacts to the aquifers above the Appin Mine.

Yours faithfully,

Neil Manewell
Technical Modelling Lead / Principal - Groundwater Modeller
Australasian Groundwater and Environmental Consultants Pty Ltd

6 References

Barnett, B, Townley, LR, Post, V, Evans, RE, Hunt, RJ, Peeters, L Richardson, S, Werner, AD, Knapton, A, & Boronkay, A (2012), *Australian groundwater modelling guidelines*. Waterlines report, National Water Commission, Canberra (herein referred to as the AGMG).

Commonwealth of Australia (CoA), (2018), *Information guidelines for proponents preparing coal seam gas and large coal mining development proposals*, Commonwealth of Australia, May 2018.

Doherty, J. and Moore C., (2021). *Decision Support Modelling Viewed through the Lens of Model Complexity. A GMDSI Monograph*. National Centre for Groundwater Research and Training, Flinders University, South Australia

Guo, H., Adhikary, D., and Gaveva, D. (2007). *Hydrogeological response to longwall mining,* ACARP Report C14033, CSIRO Exploration and Mining: Australian Coal Industry's Research Program (ACARP)

Middlemis H and Peeters LJM (2018), *Uncertainty Analysis – Guidance for groundwater modelling within a risk management framework.* A report prepared for the Independent Expert Scientific Committee on Coal Seam Gas and Large Coal Mining Development through the Department of the Environment and Energy, Commonwealth of Australia 2018.

Murray-Darling Basin Commission (2000) *Groundwater Flow Modelling Guideline*, November 2000, Aquaterra Consulting Pty Ltd, Project No. 125, Final Guideline – Issue I, 16 January 2001

Pells, S. (2017), *Groundwater Modelling of the Hume Coal Project,* https://majorprojects.planningportal.nsw.gov.au/prweb/PRRestService/mp/01/getContent?AttachRef=SSD-7172%2120190319T233221.791%20GMT

Rau,R. Acworth I, Halloran L, Timms W , Cuthbert M (2018). *Quantifying Compressible Groundwater Storage by Combining Cross-Hole Seismic Surveys and Head Response to Atmospheric Tides*

SLR (2022). *Appin Mine Extraction Plan. Groundwater Impact Assessment.* SLR. V8.0. August 2022

Murray–Darling Basin Commission (2001), *Groundwater flow modelling guideline*, report prepared by Aquaterra, January 2001

Table 1 Numerical model compliance checklist (AGMG, 2012)

Question	Comment	Yes/No
1a. Are the model objectives clearly stated?	Modelling objectives are clearly stated in Section 1.2 of the report. Model confidence level not stated. This review assessed Class 2/3 achieved (see Appendix A), which is fit for purpose.	Yes
1b. Model confidence level stated?	Missing. Based on the model report the Model confidence level is assessed as Class 2 (with some attributes of Class 3). This is not a material omission.	No
2. Are the objectives satisfied?	Model and reporting objectives are outlined in Section 1.2 are satisfied by the reported. Numerical groundwater modelling satisfies the project objectives.	Yes
3. Is the conceptual model consistent with objectives and confidence level?	An adequately detailed conceptual model is presented in the report. It reviews hydrogeology, groundwater levels, flows, surface water, hydraulic parameters. There is a detailed review and analysis of the height of fracturing and potential increases to hydraulic conductivity in the area above the longwalls.	Yes
4. Is the conceptual model based on all available data, presented clearly and reviewed by an appropriate reviewer?	Yes, the conceptual model is presented clearly and illustrated across numerous figures in Section 2 and 3.	Yes
5. Does the model design conform to best practice?	Model design detailed in Section 4. The modelling design and approach are consistent with modelling best practice. In recent times regulatory bodies have promoted the use of "stacked drains" to overcome potential disconnection through the fracture profile. The model however reproduces depressurisation signatures above the longwalls, suggested the approach is appropriate.	Yes
6. Is the model calibration satisfactory?	Section 4.2. Model calibration was carried out using PEST++. Calibration statistics are satisfactory. The model achieved good matching to groundwater level lowering in response to longwall mining. Although not presented, the model replicates longwall mining inflow adequately.	Yes
7. Are the calibrated parameter values and estimated fluxes plausible?	The history matched parameter values shown in Table 13 and 14 are plausible based on the site data and based on this reviewers experience in the region. It is unknown if the model replicates baseflow adequately.	Yes
8. Do the model predictions conform to best practice?	The calibrated model predicts groundwater drawdown, baseflow change, and estimates of inflow to the end of mine life. It is unknown how the groundwater system behaves during the recovery phase. Predictions are calculated and presented according to best practice.	Yes
9. Is the uncertainty associated with the simulations/predictions reported?	Conceptual, parametric, scenario uncertainty are not undertaken presented.	No
10. Is the model fit for purpose?	Yes, in this reviewers opinion, the model is fit for purpose for simulating and predicting groundwater inflow and potential drawdown associated with the Appin Mine Longwalls.	Yes

Table 2 Numerical model classification checklist (AGMG, 2012)

CLASS	DATA		CALIBRATION		PREDICTION		INDICATORS	
1 (Simple)	Not much/Sparse coverage		Not possible		Timeframe >> calibration		Timeframe > 10x calibration	
	No metered usage		Large error statistic		Long stress periods		Stresses > 5x calibration	
	Low resolution topo DEM		Inadequate data spread		Transient prediction but steady-state calibration		Mass balance > 1% (or one-off <5%)	
	Poor aquifer geometry		Targets incompatible with model purpose		Bad verification		Properties <> field	
	Basic/Initial conceptualisation		-		-		Poor performance stats/no review	
2 (Impact assessment)	Some data / OK coverage	✓	Weak seasonal match	✓	Predictive timeframe > calibration		Predictive timeframe = 3-10x calib.	
	Some usage info	✓	Some long-term trends wrong	~	Different stresses &/or periods		Predictive stresses = 2-5x calib.	
	Some baseflow estimates and some K/S measurements	~	Partial performance (e.g. some stats/part record/ model-measure offsets)		No verification but key simulations constrained by data	~	Mass balance < 1% (all stress periods)	
	Some high res. Topo DEM and adequate aquifer geometry	✓	Head and flux targets constrain calibration	~	Calib. & prediction consistent (transient of steady-state)	✓	Some properties <> field measurements	✓
	Sound conceptualisation, reviewed & stress-tested	✓	Non-uniqueness, sensitivity and qualitative uncertainty addressed	×	Magnitude & type of stresses outside range of cal. Stresses	✓	Some poor performance (but no coarse discretisation in key areas/times)	~
3 (Complex simulator)	Plenty of data, good coverage	~	Good performance stats	~	Timeframe ~ calibration	✓	Predictive timeframe < 3x calib.	✓
	Good metered volumes (all users)	~	Long-term trends replicated	✓	Similar stress periods	✓	Predictive stresses < 2x	✓
	Local climate data and baseflow	~	Seasonal fluctuations OK	×	Good verification or all simulations constrained by data	~	Mass balance <0.5% (all periods)	✓
	K measurements from range of tests	~	Calibration to present day data targets	~	Steady-state prediction only when calibration in steady state	NA	Properties ~ field measurements	~

CLASS	DATA		CALIBRATION		PREDICTION		INDICATORS	
	High res topo DEM in all areas & good aquifer geometry	~	Non-uniqueness minimised & or parameter identifiability/minimum error variance or RCS assessed	×	Suitable computational methods applied & parameters are consistent with conceptualisation	✓	No poor performance of coarse discretisation in key areas (grid/time)	~
	Mature conceptualisation	~	Sensitivity &/or Qualitative Uncertainty	×	Quantitative uncertainty analysis	×	Reviewed by experienced Hydro/Modeller	✓
							Criterion met at higher class	
							Criterion partially met at relevant class	~
							Criterion met at the relevant class	✓
							Criterion not met by current model study	×

Table 3 AGMG guideline model checklist

Review questions		Yes/No	Comment
1.	**Planning**		
1.1	Are the project objectives stated?	Yes	See section 1.2 of the model report
1.2	Are the model objectives stated?	Yes	Also in Section 1.2, but not segregated from the overall groundwater assessment
1.3	Is it clear how the model will contribute to meeting the project objectives?	Yes	See section 1.2
1.4	Is a groundwater model the best option to address the project and model objectives?	Yes	Assessment calls for calculating drawdown, baseflow change, inflow in a cumulative environment
1.5	Is the target model confidence-level classification stated and justified?	No	Not included
1.6	Are the planned limitations and exclusions of the model stated?	No	There is some discussion on the models inability to fit measurement data in section 4.2.2. However, there is no "Model limitations" section, or a discussion on the implied error in the model predictions.
2.	**Conceptualisation**		
2.1	Has a literature review been completed, including examination of prior investigations?	Yes	Literature review of previous reports and work at the site.
2.2	Is the aquifer system adequately described?	Yes	All aquifers are identified and justified
2.2.1	Hydrostratigraphy including aquifer type (porous, fractured rock ...)	Yes	Described in section 3
2.2.2	Lateral extent, boundaries and significant internal features such as faults and regional folds	Yes	Described in section 2
2.2.3	Aquifer geometry including layer elevations and thicknesses	Yes	Presented in Section 2.5 and Figure 7-8
2.2.4	Confined or unconfined flow and the variation of these conditions in space and time?	Yes	Described in section 3.2 and 3.6
2.3	Have data on groundwater stresses been collected and analysed?	Yes	Measured groundwater inflows discussed in Section 2. Abstraction rates from landholder wells, csg pumping are presented.

Review questions		Yes/No	Comment
2.3.1	Recharge from rainfall, irrigation, floods, lakes	No	Rainfall has been discussed in Section 2.1. There is no discussion on the range of recharge to the hydrostratigraphic units in the conceptual model section.
2.3.2	River or lake stage heights	Yes	Section 2.3 provides the flow rates and flow duration curves
2.3.3	Groundwater usage (pumping, returns etc)	Yes	local groundwater usage is described in Section 3.6 and 4.2
2.3.4	Evapotranspiration	Yes	Section 2.1 describes the evapotranspiration and potential/actual evapotranspiration estimates from BoM. However, the model uses 511 mm/year (~55% of actual).
2.3.5	Other?	NA	
2.4	Have groundwater level observations been collected and analysed?	Yes	Used as the measurement dataset for the calibration exercise.
2.4.1	Selection of representative bore hydrographs	Yes	Hydrographs are presented in Figures 11 to 30 and descriptions of groundwater behaviour in Sections 3.3
2.4.2	Comparison of hydrographs	Yes	Section 3.3 describes water levels and hydrographs are shown in Figures 11 to 30
2.4.3	Effect of stresses on hydrographs	Yes	Previous mining identified in bores screening deeper units. Climate and pumping responses discussed.
2.4.4	Watertable maps/piezometric surfaces?	Yes	Presented in Figure 23, 25, 26. Source of the data is unknown; the contour appear to have been derived from a groundwater model as opposed to an interpolated surface based on measurements.
2.4.5	If relevant, are density and barometric effects taken into account in the interpretation of groundwater head and flow data?	NA	
2.5	Have flow observations been collected and analysed?	No	No discussion on measured flow in streams/creeks or mine ingress.
2.5.1	Baseflow in rivers	No	No baseflow analysed. Main water course is gauged and stage results from this are presented.
2.5.2	Discharge in springs	NA	No springs identified in model area.

Review questions	Yes/No	Comment
2.5.3 Location of diffuse discharge areas?	Yes	Swamps identified have discussed, although they are far from the predictive impact area of the Project.
2.6 Is the measurement error or data uncertainty reported?	No	
2.6.1 Measurement error for directly measured quantities (e.g. piezometric level, concentration, flows)	No	Sources of measurement error not discussed.
2.6.2 spatial variability/heterogeneity of parameters	Yes	Variability in field measurements is discussed, the aquifer parameters are defined as uniform within geology types, however the values for same geological types varies in different model layers.
2.6.3 Interpolation algorithm(s) and uncertainty of gridded data?	No	Unknown
2.7 Have consistent data units and geometric datum been used?	Yes	It appears so, or at least the required conversions have been made as MODFLOW USG requires consistent units
2.8 Is there a clear description of the conceptual model?	Yes	See section 3.5, 3.6
2.8.1 Is there a graphical representation of the conceptual model?	Yes	See Figure 8 and Figure 33
2.8.2 Is the conceptual model based on all available, relevant data?	Yes	Conceptual model could include groundwater levels, stresses and flow directions to improve readability
2.9 Is the conceptual model consistent with the model objectives and target model confidence level classification?	Yes	
2.9.1 Are the relevant processes identified?	Yes	The key components of the bulk groundwater movement at the Appin Mine is captured in the conceptual model
2.9.2 Is justification provided for omission or simplification of processes?	Yes	Appropriate simplification to key components has taken place with evidence supporting the simplifications and omissions
2.10 Have alternative conceptual models been investigated?	No	Not initially - alternate fracturing height investigated as an alternate scenario.
3. Design and construction		
3.1 Is the design consistent with the conceptual model?	Yes	

Review questions		Yes/No	Comment
3.2	Is the choice of numerical method and software appropriate (Table 4-2)?	Yes	Modflow-USG provides a more stable numerical scheme with the control volume finite difference method over the cell centred finite difference method. Grid had been optimised to reduce cell count.
3.2.1	Are the numerical and discretisation methods appropriate?	Yes	Cell size in the mining areas ranges from 100-200m, down to 50m to represent the ventilation shafts. Pertinent creek cells have 100m resolution. This resolution is more than adequate to represent dewatering and changes to flow.
3.2.2	Is the software reputable?	Yes	MODFLOW USG is distributed by the USGS and is now the industry standard software for modelling groundwater.
3.2.3	Is the software included in the archive or are references to the software provided?	Yes	Reference provided
3.3	Are the spatial domain and discretisation appropriate?	Yes	The extent of the model is large enough to represent cumulative depressurisation. The proposed project drawdowns do not encroach on the model boundaries.
3.3.1	1D/2D/3D	Yes	3D - MODFLOW USG
3.3.2	Lateral extent	Yes	The model is bounded by no flow cells, with some CHD cells representing water storage bodies. The boundary conditions are far enough away not to influence the key predictions of the model.
3.3.3	Layer geometry?	Yes	The chosen vertical discretisation provides sufficient detail without being too simplified. The key coal seams are simulated discretely in separate model layers. The key Sandstone units are segregated according to measurement data.
3.3.4	Is the horizontal discretisation appropriate for the objectives, problem setting, conceptual model and target confidence level classification?	Yes	100m cell sizes are appropriate to meet the relevant criteria. Coarser cells may have been more appropriate to allow for faster run times, and a more robust calibration/uncertainty analysis
3.3.5	Is the vertical discretisation appropriate? Are aquitards divided in multiple layers to model time lags of propagation of responses in the vertical direction?	Yes	Model layering is sufficient to provide vertical disconnection, but still simple enough to provide a conservative response.
3.4	Are the temporal domain and discretisation appropriate?	Yes	Quarterly stress periods adequate to simulate the progression of mining and seasonality

Review questions		Yes/No	Comment
3.4.1	Steady state or transient	Yes	Both - steady state to provide initial conditions, and transient simulation that represents historical and future mining
3.4.2	Stress periods	Yes	Stress period duration is unknown between 1960 and 2009. Quarterly stress periods thereafter more than adequate to meet objectives.
3.4.3	Time steps?	No	Unknown
3.5	Are the boundary conditions plausible and sufficiently unrestrictive?	Yes	Rivers have appropriate stage heights to replicate the measured/average presence of water. Minor creeks have zero stage height. CHDs are sufficiently far from the project to not incur predictive contamination
3.5.1	Is the implementation of boundary conditions consistent with the conceptual model?	Yes	No flow boundaries are assigned where appropriate, river stage height appropriate, Water storage non-restrictive. Recharge and EVT as conceptualised.
3.5.2	Are the boundary conditions chosen to have a minimal impact on key model outcomes? How is this ascertained?	Yes	The predicted impact at steams is minimal due to the impedance of the hydrostratigraphic units. Therefore, different representations of streams would yield similar results. CHDs and No flow cells far from drawdown cone.
3.5.3	Is the calculation of diffuse recharge consistent with model objectives and confidence level?	Yes	No discussion on anticipated range of recharge. Zoned up based on outcrop geology.
3.5.4	Are lateral boundaries time-invariant?	Yes/No	CHD cells at boundary vary depending on measured water stages. No flow at boundaries elsewhere.
3.6	Are the initial conditions appropriate?	Yes	Derived from a calibrated steady state model.
3.6.1	Are the initial heads based on interpolation or on groundwater modelling?	Yes	Groundwater modelling - first stress period simulates steady state conditions, providing reliable initial conditions
3.6.2	Is the effect of initial conditions on key model outcomes assessed?	Yes	Steady state results would be affected by the changes to hydraulic conductivity and recharge explored through sensitivity/uncertainty analysis
3.6.3	How is the initial concentration of solutes obtained (when relevant)?	NA	

Review questions		Yes/No	Comment
3.7	Is the numerical solution of the model adequate?	Yes	The Overall water budget is 0.0 ML/day. It is unknown what the percent discrepancy is, or the maximum discrepancy measured at any timestep.
3.7.1	Solution method/solver	No	Not stated, although assumed that the SMS solver is used for control volume finite difference solution scheme of MODFLOW USG
3.7.2	Convergence criteria	No	Not stated
3.7.3	Numerical precision	No	Not stated
4.	**Calibration and sensitivity**		
4.1	Are all available types of observations used for calibration?	No	Calibrated to heads only. Inflows are available but only used as validation measure. No attempts at baseflow calibration in spite of numerous surface water monitoring locations near the project area.
4.1.1	Groundwater head data	Yes	Model is calibrated to water level data (up to 2021)
4.1.2	Flux observations	No	Inflow measurements used as validation only.
4.1.3	Other: environmental tracers, gradients, age, temperature, concentrations etc.	NA	
4.2	Does the calibration methodology conform to best practice?	Yes/No	Current best practice suggests all available measurements are used to form a well-posed problem for PEST to solve. Best practise also calls for a model to be endowed with many parameters to accommodate nuances in the groundwater system. Parameters that are likely to lie squarely in the solution space, namely the fracturing rate above the longwalls was not adjusted.
4.2.1	Parameterisation	Yes	Geological extents are used to define parameter zones
4.2.2	Objective function	Yes	Yes - below the suggested 10% at 4.7%
4.2.3	Identifiability of parameters	No	Composite sensitivity or identifiability of adjustable parameters not presented

Review questions	Yes/No	Comment
4.2.4 Which methodology is used for model calibration?		Calibration used PEST to perform gradient based automated optimisation.
4.3 Is a sensitivity of key model outcomes assessed against?	No	Sensitivity of the parameters to measurement data is not presented.
4.3.1 Parameters	No	Sensitivity of connective cracking multiplier was undertaken but not presented (because it failed to match historic measurements).
4.3.2 Boundary conditions	No	-
4.3.3 Initial conditions	No	-
4.3.4 Stresses	Yes/No	Tammetta fracture height explored through sensitivity. The height is lower than the Ditton approach.
4.4 Have the calibration results been adequately reported?	No	The sensitivity and range of parameters assessed during the calibration process is not reported.
4.4.1 Are there graphs showing modelled and observed hydrographs at an appropriate scale?	Yes	Some are provided in the original report - An appendix with all hydrographs used in the calibration should be presented as an appendix.
4.4.2 Is it clear whether observed or assumed vertical head gradients have been replicated by the model?	Yes	Hydrographs with measured and simulated heads are presented at several VWP locations. A chart showing pressure vs height would make the fit clearer.
4.4.3 Are calibration statistics reported and illustrated in a reasonable manner?	Yes	Scatterplot, histogram, table of residuals, SRMS/RMS presented
4.5 Are multiple methods of plotting calibration results used to highlight goodness of fit robustly? Is the model sufficiently calibrated?	Yes	Scatter diagram and hydrographs are used in conjunction with statistical measures of the error. The original 2009 calibration achieved a SRMS of 9%. Updating the model, adding more measurements, and recalibrating has reduced SRMS by approximately half.
4.5.1 Spatially	Yes	Map of average residuals presented (Figure 46)
4.5.2 Temporally	Yes	Hydrographs are shown comparing observed and simulated water levels

Review questions		Yes/No	Comment
4.6	Are the calibrated parameters plausible?	Yes	Hydraulic conductivities are typical of other assessments in the regional area. It should be noted that the Stanwell park claystone (Layer 10) is at the lower bound of conceptualised K that unit. Vertical conductivities have max kh/kv ratio of 200.
4.7	Are the water volumes and fluxes in the water balance realistic?	Yes	Volumes entering and leaving the model domain appear plausible for what they represent. Aerial recharge is generally lower than the river leakage conforming to the conceptual model where aquifers are recharged through the river. Baseflow is a large component of the budget (16%) due to the assumed incision into the aquifer.
4.8	Has the model been verified?	Yes	Model verified to mine inflow rates
5.	**Prediction**		
5.1	Are the model predictions designed in a manner that meets the model objectives?	Yes	Boundary conditions representing the dewatering from the mines are applied and impacts are defined by comparison to a null model
5.2	Is predictive uncertainty acknowledged and addressed?	No	Predictive uncertainty not undertaken
5.3	Are the assumed climatic stresses appropriate?	Yes	Quarterly averages used
5.4	Is a null scenario defined?	Yes	Null scenario includes the other mines in the model domain, but removes Appin and associated changes. A second null scenario simulates all mining with the exception of Longwalls 709-711, and 905. A third Null scenario simulates no mining in the region.
5.5	Are the scenarios defined in accordance with the model objectives and confidence level classification?	No	Model classification not discussed.
5.5.1	Are the pumping stresses similar in magnitude to those of the calibrated model? If not, is there reference to the associated reduction in model confidence?	Yes	Extraction due to mine dewatering is included in the historic period. Landholder pumping stresses continue through the predictive period.
5.5.2	Are well losses accounted for when estimating maximum pumping rates per well?	NA	

Review questions	Yes/No	Comment
5.5.3 Is the temporal scale of the predictions commensurate with the calibrated model? If not, is there reference to the associated reduction in model confidence?	Yes	Calibration period is longer than the prediction period
5.5.4 Are the assumed stresses and timescale appropriate for the stated objectives?	Yes	Quarterly stress periods adequate to simulate the progression of mining and seasonality
5.6 Do the prediction results meet the stated objectives?	Yes	Predictions show impact extent, water take from mine, and impact and landholders. "Insignificant" impact to baseflow to Creeks.
5.7 Are the components of the predicted mass balance realistic?	Yes	
5.7.1 Are the pumping rates assigned in the input files equal to the modelled pumping rates?	NA	-
5.7.2 Does predicted seepage to or from a river exceed measured or expected river flow?	No	Predicted leakage is less than baseflow. There is no discussion on measured baseflow in the creeks.
5.7.3 Are there any anomalous boundary fluxes due to superposition of head dependent sinks (e.g. evapotranspiration) on head-dependent boundary cells (Type 1 or 3 boundary conditions)?	No	No evidence of 'short circuiting' of flows between boundary conditions
5.7.4 Is diffuse recharge from rainfall smaller than rainfall?	Yes	recharge generally 0.5 to 5% of rainfall
5.7.5 Are model storage changes dominated by anomalous head increases in isolated cells that receive recharge?	No	-
5.8 Has particle tracking been considered as an alternative to solute transport modelling?	NA	-
6. Uncertainty		
6.1 Is some qualitative or quantitative measure of uncertainty associated with the prediction reported together with the prediction?	No	No discussion or simulation of parametric, structural, or conceptual uncertainty. One scenario exploring a shallower height of fracturing.
6.2 Is the model with minimum prediction-error variance chosen for each prediction?	No	Uncertainty analysis not undertaken
6.3 Are the sources of uncertainty discussed?	No	
6.3.1 Measurement of uncertainty of observations and parameters	Yes	There is some discussion on the potential for error due to structural simplification of layers; strong vertical gradients in thick layers can lead to errors of +/- 5m.

Review questions		Yes/No	Comment
6.3.2	Structural or model uncertainty	No	-
6.4	Is the approach to estimation of uncertainty described and appropriate?	No	-
6.5	Are there useful depictions of uncertainty?	No	
7.	**Solute transport**		
7.1	Has all available data on the solute distributions, sources and transport processes been collected and analysed?	NA	
7.2	Has the appropriate extent of the model domain been delineated and are the adopted solute concentration boundaries defensible?	NA	
7.3	Is the choice of numerical method and software appropriate?	NA	
7.4	Is the grid design and resolution adequate, and has the effect of the discretisation on the model outcomes been systematically evaluated?	NA	
7.5	Is there sufficient basis for the description and parameterisation of the solute transport processes?	NA	
7.6	Are the solver and its parameters appropriate for the problem under consideration?	NA	
7.7	Has the relative importance of advection, dispersion and diffusion been assessed?	NA	
7.8	Has an assessment been made of the need to consider variable density conditions?	NA	
7.9	Is the initial solute concentration distribution sufficiently well-known for transient problems and consistent with the initial conditions for head/pressure?	NA	
7.10	Is the initial solute concentration distribution stable and in equilibrium with the solute boundary conditions and stresses?	NA	
7.11	Is the calibration based on meaningful metrics?	NA	
7.12	Has the effect of spatial and temporal discretisation and solution method taken into account in the sensitivity analysis?	NA	

Review questions		Yes/No	Comment
7.13	Has the effect of flow parameters on solute concentration predictions been evaluated, or have solute concentrations been used to constrain flow parameters?	NA	
7.14	Does the uncertainty analysis consider the effect of solute transport parameter uncertainty, grid design and solver selection/settings?	NA	
7.15	Does the report address the role of geologic heterogeneity on solute concentration distributions?	NA	-
8.	**Surface water–groundwater interaction**		
8.1	Is the conceptualisation of surface water–groundwater interaction in accordance with the model objectives?	Yes	Appropriately represented such that impacts on surface water bodies can be predicted.
8.2	Is the implementation of surface water–groundwater interaction appropriate?	Yes	Impacted creeks are represented in a way that the reduction of baseflow is adequately simulated.
8.3	Is the groundwater model coupled with a surface water model?	No	Not a separate model, but the influence of the surface water system is adequately simulated by the RIV package within MODFLOW
8.3.1	Is the adopted approach appropriate?	Yes	See above
8.3.2	Have appropriate time steps and stress periods been adopted?	Yes	Stage in the higher orders show variability that is appropriately represented with quarterly stress periods.
8.3.3	Are the interface fluxes consistent between the groundwater and surface water models?	Yes	Budgets are appropriate and plausible given the incision of the creeks into the surficial aquifers.

Table 4 MDBD guideline checklist

Q	Question	Not Applicable or Unknown	Score 0	Score 1	Score 3	Score 5	Score	Max Score (0, 3, 5)
1	**THE REPORT**							
1.1	Is there a clear statement of project objectives in the modelling report?	-	Missing	Deficient	Adequate	Very Good	3	5
1.2	Is the level of model complexity clear or acknowledged?	-	Missing	No	Yes	Very Good	1	5
1.3	Is a water or mass balance reported?	-	Missing	Deficient	Adequate	Very Good	3	5
1.4	Has the modelling study satisfied project objectives?	-	Missing	Deficient	Adequate	Very Good	3	5
1.5	Are the model results of any practical use?	-	-	No	Maybe	Yes	5	5
2	**DATA ANALYSIS**							
2.1	Has hydrogeology data been collected and analysed?	-	Missing	Deficient	Adequate	Very Good	5	5
2.2	Are groundwater contours or flow directions presented?	-	Missing	Deficient	Adequate	Very Good	3	5
2.3	Have all potential recharge data been collected and analysed? (rainfall, streamflow, irrigation, floods, etc.)	-	Missing	Deficient	Adequate	Very Good	3	5
2.4	Have all potential discharge data been collected and analysed? (abstraction, evapotranspiration, drainage, spring flow, etc.)	-	Missing	Deficient	Adequate	Very Good	3	5
2.5	Have the recharge and discharge datasets been analysed for their groundwater response?	-	Missing	Deficient	Adequate	Very Good	3	5
2.6	Are groundwater hydrographs used for calibration?	-	-	No	Maybe	Yes	5	5
2.7	Have consistent data units and standard geometrical datums been used?	-	-	No	Yes	-	3	3
3	**CONCEPTUALISATION**							
3.1	Is the conceptual model consistent with project objectives and the required model complexity?	-	Unknown	No	Maybe	Yes	5	5
3.2	Is there a clear description of the conceptual model?	-	Missing	Deficient	Adequate	Very Good	3	5
3.3	Is there a graphical representation of the modeller's conceptualisation?	-	Missing	Deficient	Adequate	Very Good	3	5
3.4	Is the conceptual model unnecessarily simple or unnecessarily complex?	-	-	Yes	No		3	3

Q	Question	Not Applicable or Unknown	Score 0	Score 1	Score 3	Score 5	Score	Max Score (0, 3, 5)
4	**MODEL DESIGN**							
4.1	Is the spatial extent of the model appropriate?	-	-	No	Maybe	Yes	5	5
4.2	Are the applied boundary conditions plausible and unrestrictive?	-	Missing	Deficient	Adequate	Very Good	3	5
4.3	Is the software appropriate for the objectives of the study?	-	-	No	Maybe	Yes	5	5
5	**CALIBRATION**							
5.1	Is there sufficient evidence provided for model calibration?	-	Missing	Deficient	Adequate	Very Good	3	5
5.2	Is the model sufficiently calibrated against spatial observations?	-	Missing	Deficient	Adequate	Very Good	3	5
5.3	Is the model sufficiently calibrated against temporal observations?	-	Missing	Deficient	Adequate	Very Good	3	5
5.4	Are calibrated parameter distributions and ranges plausible?	-	Missing	No	Maybe	Yes	3	5
5.5	Does the calibration statistic satisfy agreed performance criteria?	-	Missing	Deficient	Adequate	Very Good	3	5
5.6	Are there good reasons for not meeting agreed performance criteria?	-	Missing	Deficient	Adequate	Very Good	5	5
6	**VERIFICATION**							
6.1	Is there sufficient evidence provided for model verification?	-	Missing	Deficient	Adequate	Very Good	0	5
6.2	Does the reserved dataset include stresses consistent with the prediction scenarios?	-	Unknown	No	Maybe	Yes	5	5
6.3	Are there good reasons for an unsatisfactory verification?	-	Missing	Deficient	Adequate	Very Good	0	5
7	**PREDICTION**							
7.1	Have multiple scenarios been run for climate variability?	-	Missing	Deficient	Adequate	Very Good	0	5
7.2	Have multiple scenarios been run for operational /management alternatives?	-	Missing	Deficient	Adequate	Very Good	3	5
7.3	Is the time horizon for prediction comparable with the length of the calibration / verification period?	-	Missing	No	Maybe	Yes	5	5
7.4	Are the model predictions plausible?	-		No	Maybe	Yes	5	5

Q	Question	Not Applicable or Unknown	Score 0	Score 1	Score 3	Score 5	Score	Max Score (0, 3, 5)
8	**SENSITIVITY ANALYSIS**							
8.1	Is the sensitivity analysis sufficiently intensive for key parameters?	-	Missing	Deficient	Adequate	Very Good	1	5
8.2	Are sensitivity results used to qualify the reliability of model calibration?	-	Missing	Deficient	Adequate	Very Good	0	5
8.3	Are sensitivity results used to qualify the accuracy of model prediction?	-	Missing	Deficient	Adequate	Very Good	1	5
9	**UNCERTAINTY ANALYSIS**							
9.1	If required by the project brief, is uncertainty quantified in any way?	-	Missing	No	Maybe	Yes	1	5
						TOTAL SCORE	**108 (61%)**	**181**