24 September 2018
WRL Ref: WRL2017018 L20180924

Mr. Clay Preshaw
Department of Planning and Environment
320 Pitt Street Sydney 2000

cc: Peter Martin, Alan Lindsay
Coal Free Southern Highlands Inc.

UNSW SYDNEY

Water Research Laboratory

Dear Clay,

WRL Response to Hume Coal 'Response to Submissions'

The Water Research Laboratory (WRL) of the School of Civil and Environmental Engineering at UNSW Sydney has reviewed the Response to Submissions (RTS) document tendered to NSW Department of Planning and Environment (NSW DPE) in support of the Hume Coal Prospect and provides this response. The review was funded by the Coal Free Southern Highlands Inc. The response was prepared by WRL Principal Engineer for Groundwater and Modelling, Mr Doug Anderson. Mr Anderson's CV is available online at: http://www.wrl.unsw.edu.au/staff/doug-anderson.

The response is provided as an Appendix which is ordered as follows:

Thank you for considering Mr Anderson's comments on this project. If the Department has any questions regarding this review please contact Mr Anderson in the first instance (email: d.anderson@wrl.unsw.edu.au).

Yours sincerely,

Grantley Smith
Manager

Water Research Laboratory
School of Civil and Environmental Engineering | UNSW SYDNEY
110 KING ST, MANLY VALE, NSW, 2093, AUSTRALIA
T +61 (2) 8071 9800 | F +61 (2) 9949 4188 | ABN 57 195 873 179 | www.wrl.unsw.edu.au
Quality System Certified to AS/NZS ISO 9001
Innovative answers for tomorrow's water engineering questions, today | Since 1959

GREEN GLOBE Winner 2015

1. Key concepts and background to this response

Within a scientific framework, the purpose of groundwater assessment is to understand and manage subsurface groundwater as best possible. This is achieved by hypothesising how it functions and then building models (e.g. conceptual and numerical models) to test all the assumptions made in developing that understanding prior to providing prediction. Hypothesis testing therefore tests assumptions to reject false hypotheses and to understand assumptions made to address knowledge gaps. Therefore, groundwater modelling is just as much about hypothesis testing to ensure the job is done properly as it is about providing calibration and prediction by trial and error or with a piece of software that automates the selection of model inputs (often with minimal constraint) to match hydraulic head (and flow) observations.

In undertaking such work it must be appreciated that the time scales of mining projects are long and mining impacts are even longer; sometimes longer than the professional lives or availability of the professionals who undertook the work. Therefore, at least in the scientific literature, it is readily accepted that the critical focus of model analysis, and therefore the review of any model, is to ensure that all models (geological, conceptual, numerical) are well documented, appropriate and that the numerical model makes unbiased, best possible use of the available field data.

In 2005 the internationally respected hydrogeologist John Bredehoeft (2005) stated:

> "*The foundation of model analysis is the conceptual model. Surprise is defined as new data that renders the prevailing conceptual model invalid; as defined here it represents a paradigm shift. Limited empirical data indicate that surprise occur in 20-30% of model analyses. These data suggest that groundwater analysts have difficulty selecting the appropriate conceptual model.*"

Pells and Pells (2009) highlight the importance of documenting heuristics in models which they define by quoting Taleb (2012) as:

> "*…simplified rules of thumb that make things simple and easy to implement. Their main advantage is that the user knows they are not perfect, just expedient, and is therefore less fooled by their powers. They become dangerous when we forget that.*"

The importance of clearly describing and documenting in reporting what constitutes the mediums through which, and how much, water is conveyed and which parts of the ground contain water that is under pressure (confined) or not (unconfined) is also critical. Frequently reports are ambiguous in this regard. Pells and Pells (2009) highlight this point by quoting from the final paper of one of the founding fathers of hydrogeology, C V Theis who wrote a famous paper on groundwater flow in 1935. In his final paper, "*Aquifers, Ground-Water Bodies and Hydrophers*", Theis (1987) stated:

> "*Thus, "aquifer" has been used in so many different senses by so many people to express their own particular ideas that it has become an Alice-in-Wonderland word that means just what the author says it means. Worst of all, the author practically never tells us what he means. It has been used in so many different ways that it must be abandoned entirely as a scientific word or alternately to express only the original usage of it without any relation to the water table…*"

In the opinion of WRL's reviewer, generating reliable, unbiased predictions of future groundwater impact requires model analysis based on the scientific method and an appropriate documentation. Appropriate model analysis comprises of two fundamental steps and various sub-steps that include:

1. Conceptualisation:
    a. Collecting sufficient geological, geophysical, hydrogeological and hydrological field data;
    b. Analysing this data to understand the hydrogeological structures and parameters and their uncertainty, how groundwater flows in response to the understanding and how it might flow based on the gaps in knowledge;

    c. Clearly documenting through what (and how much) groundwater flows;
    d. Clearly documenting where and when that flow is confined and unconfined; and
    e. Representing this understanding in an unbiased manner in a conceptual model.
2. Simulation:
    a. Translating this conceptual understanding into simulators of groundwater flow;
    b. Test key assumptions (hypothesis testing);
    c. Running simulators to generate predictions of future impacts with sensitivity analysis; and
    d. Running uncertainty analysis to address uncertainty in (1).

On 23 June 2017 WRL submitted comments to NSW DPE (our ref: WRL2017018DJA L20170623) drawing attention to potential issues and various gaps in conceptualisation and simulation with regard to the Hume Coal EIS groundwater impact assessment. While various concerns were raised regarding simulation (including a lack of uncertainty analysis and the representativeness of the boundary conditions that temporarily allow groundwater to enter the mine workings / voids) a fundamental criticism that was levaged against the assessment was various shortcomings in conceptualisation, documentation, and simplifications / assumptions in modelling that ignored field data and knowledge to generate predictive bias (which was not then assessed with hypothesis testing to understand the implications of this), and uncertainty analysis of knowledge gaps.

In the opinion and experience of WRL's reviewer such shortcomings in conceptualisation are critical and leave the door open for environmental and financial outcomes in the distant future to be substantially different from prediction; an outcome known to all practicising hydrogeologists light-heartedly and perhaps inappropriately after Bredehoeft (2005) as '*conceptual model surprise*'.

In the experience of WRL's reviewer, his review of annual reporting for coal mines finds a number of operational underground mines in the Sydney Coalfields that routinely need to re-calibrate their predictive models because of observation deviating substantially away from prediction because of simplification employed at the EIS stage. In some cases, there is even concern amongst the community, media and even NSW DPE that some mines have caused environmental impacts not predicted during environmental impact assessment (e.g. http://www.abc.net.au/news/2018-09-17/sydney-coal-mine-ordered-to-repair-cracked-creek/10253148).

2. Summary response to the Hume Coal RTS document

In the opinion of WRL's reviewer, the RTS does not address a number of the issues and advices provided previously by WRL. In particular, it remains WRL's reviewer's opinion that the EIS and the RTS has not demonstrated appropriate and unbiased conceptualisation of geological structure and hydrogeological properties both on paper and in the numerical simulator base case.

Therefore, WRL's reviewer remains concerned that bias has been translated into the model predictions and (now also) into the uncertainty analysis provided within the RTS. This review also finds that the 'post-processing' of the uncertainty analysis is deficient (refer Section 3.4). This is concerning because valid uncertainty analysis and hypothesis testing of assumptions is a critical component of modelling.

To restate the key concerns in WRL's original submission, and now again for the RTS, Hume Coal and their consultants have not:

1. Applied the scientific method (e.g. presentations of data, scientific data analyses, and interpretations of results) to reject the concerns and scientific hypothesis provided by project commentators in regards to the EIS model conceptualisation (e.g. geological structure and hydrogeological parameters). Demonstration of appropriate technical analysis via reporting is

negligible and, other than the new uncertainty analysis, the RTS is largely based on expert opinions, from practitioners with field experience at different sites.

2. Adequately addressed the EIS submissions on the following three key issues:
    a. Appropriate justification for the simplistic subsurface 3D geological conceptualisation via hypothesis testing within the numerical model tool utilising the collected field data;
    b. Demonstration that field measurements of hydrogeological properties were assigned to the 3D geological framework in an unbiased manner; and
    c. Demonstration that appropriate boundary conditions were utilised for numerical model conceptualisation, calibration and prediction.
3. Applied rational cut-offs to their uncertainty analysis predictions.

Section 3 re-describes key concerns of WRL's reviewer in response to the material presented in the RTS. Additional additional issues are discussed in Section 4.

3. Key Issues

3.1 Hypothesis testing

While the RTS is critical of the work completed by Pells Consulting (e.g. Volume 1, Pages 180 and 186) it must be recognised that the Southern Highlands Community commissioned Pells Consulting to develop a simple groundwater flow model of the Hume Coal Prospect based on publically available knowledge and data in which the assumptions relied upon by Hume Coal in their groundwater impact assessments could be tested, the sensitivities reported and assumptions questioned. WRL reviewed this numerical groundwater model twice in accordance with the Australian Groundwater Modelling Guidelines and assessed it was fit for this purpose (our refs: WRL2013073DJA L20131211, WRL2017018DJA L20170623).

Consequently, WRL's reviewer has carefully reviewed the criticisms made against the Pells Consulting conceptual and numerical models in the RTS. The conclusion of this review is that Hume Coal and their experts may have missed the point of the submissions by Pells Consulting.

One of the key points illustrated by Pells and Pells (2013) and Pells and Pan (2017) is that the manner in which the geology is simplified has a substantial bearing on the prediction of impact. Hume Coal simplified the geology to one extreme and the community perhaps simplified to the other extreme. WRL's reviewer does not know which is more correct because the geological model and data is not publically available nor apparently available to NSW Government DI Water for their assessment. Therefore the only comment that WRL's reviewer can provide is that after three (3) groundwater modelling iterations, Hume Coal have not utilised their model or data to test the validity of their simplification hypothesis to represent the geology and hydrogeology in the way they describe.

Therefore, WRL's reviewer would welcome a technical report and numerical model assessment that incorporates an assessment of all the geological, geophysical and hydrogeological data and hypothesis testing that was completed to demonstrate that the Hume Coal groundwater impact assessment is sound. Without such reporting NSW Government will be unable to understand the inherent heuristic biases in the Hume Coal EIS.

3.2 Geological model simplification bias

The Hawkesbury Sandstone Formation was deposited over a period of 200 million years. Pells, P. (1993) described the general layered nature of the Hawksbury Sandstone Formation, which comprises sheet sandstone, massive sandstone and mudstone facies, during the EH Davis Memorial Lecture of 1993 (http://www.pellsconsulting.com.au/selectedPapers/hawkesburySandstone.htm). Some photos that highlight some of the age (i.e. depth) dependent layers in the Hakwesbury Sandstone can be viewed on Wikipedia (https://en.wikipedia.org/wiki/Sydney_sandstone). Pells (1993) describe "*most units of the mudstone exhibit a fairly uniform thickness and appear to be sheet-like, although laterally discontinuous and frequently terminated laterally by erosion surface overlain by massive facies sandstone. This sudden lateral termination can make borehole interpolation very difficult*".

The RTS contends that the Pells and Pells (2013) and Pells and Pan (2017) "*sub-division of the Hawkesbury Sandstone into three consistent layers (A, B, C after Lee (2009)) across the whole Southern Highlands area is problematic as subtle facies changes are common for this braided stream deposit, and there is variable cementation of the bedding within the sandstone…*". Given this complexity, and the wording elsewhere in the RTS, WRL's reviewer interprets that this complex Hawkesbury Sandstone Formation is represented in the EIS base case model as a homogenous slab of rock with constant layer wide hydraulic properties decreased uniformly with depth in each model layer.

In effect the modellers appear to have simplified the complexity of comprehensive interpretation of borehole data away by their interpretation of average rock properties from the project site and other sites across the Sydney Basin, hoping on average this provides a reasonable basis for the geology above the proposed mining panels. Reliance on distant observation and averages could be considered a reasonable start if no site data was collected throughout the mine lease, however, Hume Coal argue in their RTS that they have collected an extensive field data-set.

Therefore, given the resources spent on sourcing field data, it is of concern to WRL's reviewer that this data was then utilised only subjectively in site conceptualisation to make limited cross-section figures (e.g. Figure 3.1 of the RTS) and not scientifically incorporated into the numerical model structure or the calibration objective (error minimisation) function. These functions ensure subjective bias error in a resultant model calibration is minimised and all observation data is given appropriate 'weight'.

Specifically, the EIS and RTS provides no demonstration of having undertaken any of the following work to make best use of of the field measurement data that was collected:

1. Correlate detailed geological and geophysical logs and interpolate this detailed metre by metre data across 2D cross-sections and then in 3D for the study domain;
2. Conduct spatial data correlation and rescaling work to assign the available hydraulic property data (hydraulic conductivity, specific storage and specific yield) to the known lithology;
3. Conduct upscaling exercises to compute the bulk average hydraulic properties in each computational numerical model layer and then use these for simulation; and/or
4. Develop objective functions for model calibration that apply penalties to model solutions that utilise values of hydraulic conductivity, specific yield, specific storage and recharge that deviate away from the measurements.

Therefore, WRL's reviewer advises that Hume Coal has not presented sufficient hypothesis testing of geological, geophysical and hydrogeological data and its simplification in numerical modelling to demonstrate that their model predictions are not biased by their simple conceptualisation of the Hawkesbury Sandstone Formation, i.e. as one uniform slab of rock without any age (i.e. depth) or spatially dependent variation in hydraulic properties (other than a linear decrease with depth which ignores the presence of mudstone beds that control groundwater flow).

3.3 Predictive bias and make-good arrangements

WRL's reviewer also advises that the act of simplifying the local variations in sheet, massive and mudstone facies and rock defects that do support discrete aquifers is problematic for another reason. If actual hydraulic conductivity and specific storage measurements about landholder wells from pump test analysis or data are ignored and replaced with different values in the predictive model, as has occurred in the Hume Coal assessment, then this may have significant implications for the accuracy of modelled predictions of drawdown at landholder wells.

In hydroscience this is known as the Representative Elementary Volume (REV) problem. If the model grid cells represent the average of the hydraulic properties in that space, then if enough wells are drilled into that grid cell and the model has been well calibrated without bias, then eventually the average of the available observations will match the prediction provided by the model. However, if the model uses a completely different value for the sake of heuristics (parsimony in modelling) then there will always be mismatch between model and observation.

It is also important to note the model residuals reported in the EIS and RTS provide an indication of only the best case reliability of the model to predict water levels at any point in space prior to the commencement of mining activity. As the mining progresses, wherever there is bias, the model will become increasingly inaccurate. This is discussed further in Section 3.6.

In this context, WRL's reviewer has noted that the predictions of Hume Coal's model are proposed to be a reference point for what does and does not constitute a mining impact at a landowner well. Therefore, WRL's reviewer was surprised to read that the EIS and RTS does not explore the impacts of their simplification bias as discussed throughout this response. In the opinion of WRL's reviewer, the proponent should be required to provide this assessment. Rau et al (2018) presents a nomogram which can be used to calculate the implications of mismatches in storage coefficients on drawdown predictions at irrigation wells in the context of aquifers with hydraulic conductivity ranging from 0.1 to 10 m/day.

Without such an assessment it would be rather difficult to agree on what drawdown constitutes a mining impact and what drawdown constitutes irrigation activity. Therefore, when impacts do occur this could result in costly litigation for all parties involved unless Hume Coal accepted responsibility for all drawdown irrespective of it's true source.

3.4 Truncation of uncertainty analysis results

Any model that contains bias in its conceptualisation or simulation base case will transmit that bias into the model uncertainty analysis. All models contain some degree of bias; it is simply unavoidable during simplification of a complex reality where measurement data are limited. Therefore, in the opinion of WRL's reviewer, it is exceedingly unwise to accept the application of any filters that exclude the results of uncertainty analyses model runs from reporting as this may just serve to elimate the few model predictions that are actually realistic.

In the opinion of WRL's reviewer, one arbitrary and irrational filter applied by some consultants is the the 33-67 percentile (one standard deviation) of all monte-carlo model runs 'rule'. For this rule to be applied wisely the sum of all available field observations, data analysis, model conceptualisation and simplification in a numerical simulator must result in a modelling tool base case that deviates from various measures of reality by no more than 17% (67% - 50%). Since the accuracy of a model can not possibly be known in advance (hence the reason for undertaking an uncertainty analysis), and especially if hypothesis testing of assumptions has not been demonstrated, truncation of results may result in the true prediction being excluded from consideration as is illustrated in Figure 1.

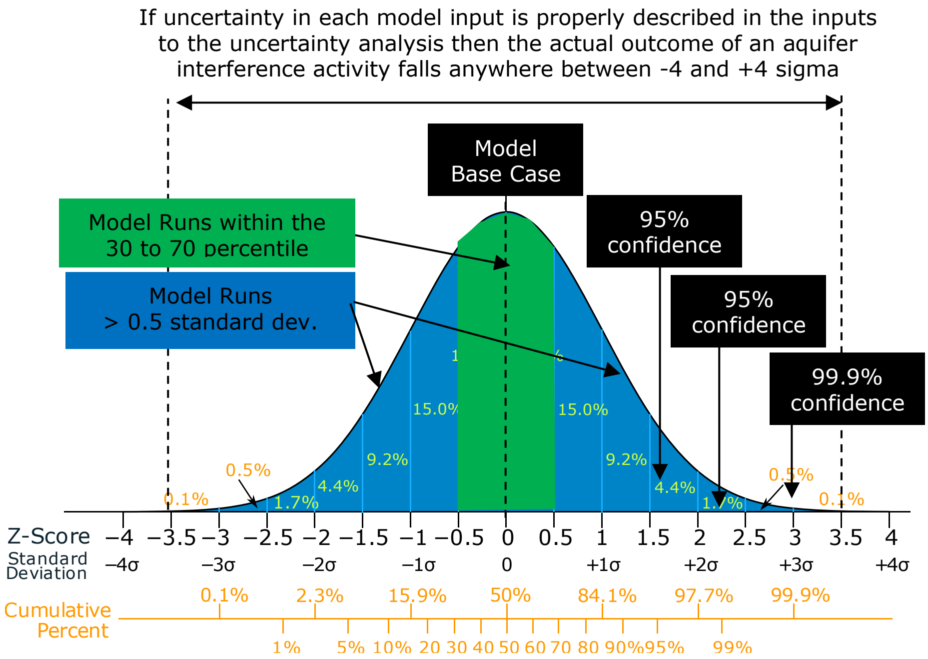

Figure 1: Bias in model calibration and potential implications for uncertainty analysis

Therefore, while some modellers may contend that the sum of all available measurements and model conceptualisation does result in a prediction that is within 17% of reality, this would appear to be in conflict with the peer reviewed advice of Bredehoeft (2005). It must also be recognised that most groundwater models are constructed from data collected at just a handful of boreholes that usually measure much less than 0.001% of the earth's surface within the study domain and therefore rely on extensive interpolation and extrapolation. Furthermore these models are often based on time series observation data that have limitations in coverage or quality and which are frequently not analysed to the fullest extent possible which can lead to substantial model bias (Rau et al., 2018). Most numerical models also employ great simplification in spatial distributions of geology and hydraulic parameters and temporal variability in boundary conditions that may allow one variable (e.g. mass balance) to be well predicted but another not so well (e.g. drawdown). Therefore, again, truncating uncertainty analysis results is very unwise.

Indeed, across all other scientific disciplines when making limited measurement and prediction of some population based on a single variable it is both normal, customary and accepted to consider either the 90th, 95th or 99th percentile confidence intervals depending on the level of risk.

Therefore, given that measurement errors are additive, and models may be based on substantial measurement and/or simplification, interpolation and extrapolation errors that introduce bias, and especially then without any reporting of hypotheses testing of key assumptions, WRL's reviewer recommends that NSW DPE consider the 5th to 95th percentile predictions of impact reported by the proponent for the Hume Coal Prospect and not the 33rd to 67th percentile.

If there is any concern about the impartiality of this advice, WRL recommends that NSW DPE retain a respected statistician who is more qualified to comment on matters of statistics and uncertainty.

3.5 Draft IESC advice on uncertainty assessment

The draft IESC advice on uncertainty assessment (Middlemiss and Peeters, 2018) states on page 13, Section 8.1 (blue text is WRL emphasis):

> *The level of hydrogeologic complexity incorporated in any model should be commensurate with its purpose (Neuman and Wierenga, 2003). It is worth reiterating for the purpose of this Explanatory Note, the purpose of a modelling study is to provide information about the uncertainty in conceptualisations and model simulation outputs in a way that allows decision makers to understand the effects of uncertainty on project objectives and the effects of potential bias.*
>
> *Refsgaard et al. (2012) concluded that the importance of geological models is lessened for flow modelling simulations, provided that (history match) conditional calibration against head and discharge data is performed, and that model simulations are confined; (i) to the same types of variables used for conditional calibration (e.g. head and flux data), and (ii) to similar hydrological stress regimes (pumping, climate and timeframes). These principles are consistent with the AGMG guiding principles (Barnett et al. 2012). It is argued that, in these cases, the inevitable (unknown) errors in the geological interpretations can (to some extent) be compensated by the biased parameter values of the history-match model. However, they warn that geological model uncertainties become crucial in situations where groundwater models that are history-matched to head and discharge data for the historical pumping or climate record are then used for extrapolation beyond that conditional calibration base. In such 'out of range' simulations, the geological structure uncertainty may often be the dominant source, and thus alternative hydrogeological conceptualisations should form part of the uncertainty assessment.*

Noting that the proposed Hume Coal project will generate the largest drawdowns and groundwater discharges ever experienced within the project footprint and immediate surrounds, and it is widely accepted that the Hawkesbury Sandstone Formation in the Southern Highlands supports highly productive aquifers (Pritchard, 2004), we interpret from the IESC guidance that the text highlighted in blue applies, i.e. that "*alternative hydrogeological conceptualisations should form part of the uncertainty assessment*". The need for specific hypothesis testing of geological complexity and incorporation into the model calibration objective function is further substantiated in the context of the Hume Coal RTS below.

3.6 Hydraulic Diffusivity and Model Calibration

Hydraulic diffusivity is a numerical model input that is defined by the ratio of hydraulic conductivity to specific storage. Hydraulic diffusivity determines the speed at which hydraulic head disturbances from mining propagate through confined aquifers to impact distant pumping wells and surface water bodies.

If a model is based on realistic, measured values of hydraulic diffusivity it is well calibrated for transient prediction. If these measurements do not exist within and around the area to be mined, then there is limited confidence that the model will honour subsequent field observations. Similiarly, if data is available (e.g. pump test flows and water level observations) but this has not been utilised in the model as part of the calibration objective (error minimisation) function because the model time-step or grid size is too coarse, then the confidence in the model prediction is also low. In such cases hypothesis testing is required and uncertainty analyses must be comprehensive, not selective.

For those modellers concerned about how best to represent aquifer test data in large regional models, there are published sub-domain modelling techniques (analytical and numerical) that allow aquifer test data to be 'nested' into regional model grids without compromising model run-time. Such techniques are best practice and are routinely used by some overseas consulting firms but were not used as part of the Hume Coal RTS.

3.7 Justifications for Confined Aquifer Storage Coefficients (Specific Storage)

In response to review provided by Pells and Pan (2017), the RTS describes revising the specific storage in the EIS from values chosen through 'calibration' to values consistent with values determined through interpretation of aquifer pumping test data. This may be sensible, however, utilising specific storage values from aquifer test interpretation without also adopting the same hydraulic conductivity values is problematic. This is because the hydraulic diffusivity of the aquifer (Section 3.1) is then misrepresented.

WRL's reviewer notes that Hume Coal do not publish the pumping test interpretation data-sheets with their analyses of hydraulic conductivity and specific storage that were utilised to support the assessment. Differences between hydraulic diffusivity observed in the field and within their model are not discussed.

Presumably as part of the geotechnical work completed for the mine there would be a wealth of rock strength and geotechnical laboratory test data and also time-series hydraulic head observation data from which specific storage values could be calculated. The EIS and RTS does not describe undertaking such technical work.

3.8 Transient calibration to aquifer test data

This issue (Section 3.7) is raised because on page 26, Volume 2D, Table 11, the RTS reports the model calibration is essentially insensitive to a four-fold increase in specific storage. This is concerning because a four-fold increase in specific storage must result in a four-fold reduction in the rate at which pressure disturbances travel through the groundwater system. Therefore, this must cause less drawdown at groundwater wells surrounding the mine (because more water is captured from storage closer to the mine).

Review of the EIS and revised RTS model predictions does indeed highlight that predicted impacts of the mine have been decreased by the most recent changes to the model. Therefore the model <u>is sensitive</u> to the specific storage parameter. Consequently, in the opinion of WRL's reviewer, the statement in the RTS that the transient calibration is insensitive to specific storage only serves to highlight 'the worth' of Hume Coal's transient calibration and the potential dangers of interpreting certain adjectives in their report too broadly.

WRL's reviewer therefore advises that the EIS and RTS transient model calibration contains no data that 'productively' informs either the aquifer specific storage or hydraulic diffusivity parameters of the model. Examination of the EIS and RTS highlights that the time-steps for simulation were approximately 180 - 182 days. This is far too coarse to allow for any formal pumping test, seasonal irrigation activity or recharge pulses to be input into the model or utilised successfully in a calibration objective function. Therefore it can be stated with confidence that the Hume Coal groundwater model demonstrates no meaningful calibration to transient aquifer test data, irrigation activity or recharge.

Meaningful transient calibration must have correspondence to the large gradients in hydraulic head that will be created by mining of the coal seams and the filling of the voids with water (not the low hydraulic gradients encountered under standard environmental conditions). Page 21 of the Australian Groundwater Modelling Guidelines (Barnett et al, 2012) clearly states that for a Class 1 model a key indicator of "calibration" is that "*stresses in prediction are more than 5 times than those in calibration*". Therefore the Hume Coal Model does not achieve this for aquifer test data and is Class 1 in this respect.

3.9 Transient 'Calibration' to Berrima Mine

Hume Coal assert that their model is calibrated to transient observation data obtained at nearby Berrima Mine. Berrima Mine was a high hydraulic-gradient aquifer interference aquifer activity when mining started, however, hydraulic heads, gradients and flows have declined since mining started. WRL's reviewer notes that neither the EIS or RTS provides, as a calibration metric, a time-series plot of the mine inflow and hydraulic head observation data from around Berrima Mine from the beginning of the life of that mine to present nor the corresponding model predictions that match those declines both for groundwater levels and flows.

As per WRL's original submission one modeller of Berrima Mine, Katarina David (pers comm. 21 June 2017) stated "*The only very accurate information at Berrima are the inflows and wq at the discharge point , all else is very limited, including geology, bores and water levels*". Therefore, if the time history of water levels around Berrima Mine are largely unknown, there must be limited scientific confidence, with reference to the transient equation for groundwater flow, that the hydraulic conductivity and specific storage values in Hume Coal model reproduce both observed flows and unobserved hydraulic heads. Since a good quality transient calibration requires access to both flow data and associated water level responses (Middlemiss and Peeters, 2018), it is the opinion of WRL's reviewer that the Hume Coal model cannot be 'calibrated to Berrima Mine' because the pre-requisite observation data does not exist.

Therefore while it can be suggested that calibration has been attempted based on only observed mine inflows at Berrima, it must be recognised that a calibration attempt based on late-time outflows from those voids, without early time inflow or hydraulic head observation data, is highly dependent on the modellers' assumption(s) of the initial groundwater levels at Berrima mine. From the RTS, WRL's reviewer understands this assumption was not tested in calibration and is therefore another point of uncertainty that was not addressed through uncertainty analysis.

4. Other issues

4.1 Interpretation of laboratory permeability tests

The RTS attributes the following statement to the NSW DPE reviewer:

> "*The modelled Hawkesbury Sandstone Kh values are reasonable in that they lie in the middle of the range of observed values (Figure 9.7); clearly not at the high end of the range as suggested by Pells [and Pan] (2017), but also not at the low end of values (mainly from core testing, indicated by grey dots in Figure 9.7)*" – RTS Volume 1 page 186.

It must be clarified that this statement is overly simplistic and potentially misleading.

Firstly, it is not appropriate or accurate to directly compare bulk-average, aquifer-scale values of horizontal hydraulic conductivity determined through aquifer test analysis to laboratory measurements on tiny pieces of drill core which are typically (more often than not) measurements of vertical hydraulic conductivity). Secondly Figure 9.7 does not distinguish which laboratory samples are measurements of vertical and hydraulic conductivity, nor which types of lithology (e.g. shale, mudstone, massive sandstone, sheet sandstone) each measurement applies to. Thirdly the graph does not indicate which of these samples were tested unloaded and which were reloaded to the correct insitu stress level. Measurements that are not made at the reloaded stress level of the aquifer can be substantially biased (e.g. Domenico and Mifflin, 1965).

Therefore, in the opinion of WRL's reviewer, any direct comparison of aquifer-scale hydraulic conductivity measurements to unspecified laboratory core measurements is potentially misleading.

4.2 Method for incorporating drill core hydraulic data into models

The ideal approach to make use of drill core hydraulic conductivity data for high risk projects is to assign the laboratory measurements to a detailed lithological description, then to examine the detailed geological and geophysical log data for that core hole and then to compute bulk-average values of horizontal and vertical hydraulic conductivity in each of the model layers by interpolation of available data using the geometric and harmonic means respectively. Then once these model-layer scaled estimates are available they need to be upscaled further (increased in hydraulic conductivity) before being compared to the aquifer-scale measurement data. This increase is needed because small core samples of 63 or 84 mm diameter rarely contain the permeable rock defects and structures observed at the scale of say 100 m x 100 m (the size of a typical numerical model grid cell). It is well established in the scientific literature, that defects and structures at scales larger than a piece of drill core dominate the majority of flow and transport in most hard rock groundwater systems (fractured drill core can rarely be tested in the laboratory).

Therefore, WRL's reviewer adopts a contrary position to the RTS and the DPE reviewer to state that the bulk-average values of horizontal and vertical hydraulic conductivity determined from aquifer tests are in nearly all circumstances the most appropriate measurements for direct input as initial estimates into regional groundwater flow models.

An analogy from which to understand this argument is to consider, and this is an example only, that on average perhaps 70% of groundwater flows through the defects in the rock mass quicikly and 30% flows through the pore space in the rock mass slowly. Therefore the majority of the water that flows into mine workings comes from groundwater flow through the rock defects. Therefore if the defect spacing in the Hawkesbury Sandstone Formation is assumed to be reasonably consistent on the scale of say a 100 m x 100 m model grid cell then the hydraulic conductivity in the model grid cell should ideally be set very close to the hydraulic conductivity observed in a groundwater well that intersects those defects. The corollary of this analogy is that there is little point setting the hydraulic properties in the model grid cell to those of the rock matrix (obtained from a well that does not intersect any fractures) if the matrix only supply 30% of the groundwater that flows into the mine workings.

WRL's reviewer understands that the EIS and RTS provide no assessment of the proportion of groundwater that flows through defects and the rock matrix. This would be an informative assessment for substantiating the values of hydraulic conductivity utilised within the EIS and RTS models and the likely impacts of the development.

4.3 Representativeness of aquifer test data

Interpretation of aquifer test data typically yields different values of hydraulic conductivity data depending on the length of the test and the time of interpretation. At early time the interpreted hydraulic conductivity values are most likely to be representative of defects, especially those in close proximity to the well. At late time the interpreted hydraulic conductivity values are most likely to be representative of some reasonable average of both the hydraulic conductivity of the rock matrix and the defects.

Therefore, if pump tests are undertaken for long enough, the interpreted late-time values of hydraulic conductivity from aquifer tests are the appropriate values to use as the basis for simplistic regional groundwater flow models. WRL's reviewer understands that the EIS and RTS considers that aquifer pump tests were undertaken for a sufficient duration of time and therefore he has difficulty accepting the logic provided in the RTS in support of homogenous hydraulic properties in individual model layers and some of the low hydraulic conductivity values that are utilised.

In the case of landowner wells intersecting regionally significant structures that provide high bore yields, it is imperative that such structures be incorporated into the model as these structures will dominate the drawdown outcomes at the landowner wells. The implication of ignoring this, which was discussed in Section 3.3, is that the model will not reliably predict the drawdown at landowner wells.

In the opinion of WRL's reviewer, a groundwater impact assessment, especially one that demonstrates exceedance of the minimal impact criteria of the NSW Aquifer Interference Policy, requires comprehensive assessment of the drawdown likely to be experienced at a landowner's well. In the opinion of WRL's reviewer, this can only be achieved by parameterising the model with localised variations to the hydraulic conductivity, specific storage and specific yield around the landowner wells to honour the available aquifer test data. Therefore, WRL recommends that all wells to be substantially impacted by the proposed development must be pump tested (if not already) with suitable observation bore networks in place to ensure the model provides accurate representation of the hydraulic properties to support the predictions that might be proposed for use as make-good arrangements.

The RTS argues:

> "*Hydraulic conductivity values adopted by Pells and Pells (2013) are not supported by regional data or sufficient pumping test analyses... There is no evidence of a consistent high hydraulic conductivity layer in the lower Hawkesbury Sandstone from the field investigations, and therefore there is no justification for inclusion of such in the Hume Coal model.*" – RTS Volume 1, Page 186

The response of WRL's reviewer to this argument, in the context of the Hume Coal model, was provided in Sections 3.6 to 3.9.

The RTS also argues that the Pells consulting models are mistaken in their interpretation of hydrogeology because:

> *"Yields from water bores that intercept deeper fractures can be very high yielding across localised areas of the Southern Highlands (URS Australia 2007 and Parsons Brinckerhoff 2009b). With the added advantage of significant available head, deep bores can pump high volumes (up to 50 L/s) at selected locations. This leads to the Hawkesbury Sandstone appearing to have higher hydraulic conductivity at depth however, this trend is more related to (localised) secondary permeability features such as fractures rather than high (regional) primary permeability."*

The response of WRL's reviewer to this argument is provided in Section 4.3 paragraph 4 above and in Section 3.7.

4.4 Confidence level classification – Pells and Pan Models

In the opinion of WRL's reviewer, the RTS is transfixed by matters of model confidence level classification and that this becomes a distraction from the key scientific shortcomings of their work as described above. As stated in Section 3.1 the Pells Consulting models were created for hypothesis testing of assumptions, something that Hume Coal appear to have overlooked despite developing three groundwater models.

In the opinion of WRL's reviewer, it is the Hume Coal model, not the Pells Consulting model, that is supporting a State Significant Development application to interfere with a highly valued aquifer. Therefore, the geological modelling, hypothesis testing and uncertainty testing within it should be better than that in a low cost model developed by the residents of the Southern Highlands. As the previous eleven (11) pages elaborate, whether the EIS and RTS model achieve this outcome is a highly debatable topic.

4.5 Confidence level classification – Hume Coal Model

Page 19, first paragraph, last sentence of the Australian Groundwater Modelling Guidelines (Barnett et al., 2012) states "*If a model falls into a Class 1 classification for either the data, calibration or prediction sectors, it should be given a Class 1 model, irrespective of all other ratings*". Therefore, utilising the example in Section 3.8, if a model is not calibrated to the key hydraulic diffusivity and specific storage parameters that control transient impacts it must be a Class 1 model.

Some practioners dislike this harsh classification scheme and attempt to dismiss Page 19, first paragraph, last sentence of the guidelines, however, there are good reasons why the guidelines were written like this (e.g. as explained for the case of hydraulic gradients at Section 3.3 and also by the IESC as reproduced in Section 3.5).

For example from the RTS, we understand the NSW DPE Reviewer is attributed to have stated that Page 19, first paragraph, last sentence of the guidelines is a mistake by the authors of the guidelines. However, the RTS does not substantiate this with any correspondence from the authors of the guidelines (Barnett et al., 2012).

The argument given in the RTS in support of the view that the guidelines are "internally inconsistent" is to make reference to a sentence elsewhere in the guidelines which indicates (and WRL's reviewer summarises) *if part of a model is in Class 1, then that model is not Class 3.* In the opinion of WRL's reviewer, that sentence in the guidelines is an example of deductive (and redundant) logic to exemplify the statement made in the guideline at Page 19, first paragraph, last sentence. Therefore it does not allow Hume Coal to upgrade their model classification to Class 2.

Therefore, in the opinion of WRL's reviewer the subsequent statements provided in the RTS - and attributed to the NSW DPE reviewer – do not withstand scrutiny via the science of critical reasoning. For example, subsequently in the RTS it is argued that NSW Government DI Water "cherry-picked" "internally inconsistent" errors from the guidelines and therefore a number of their concerns reported to NSW DPE do not apply. WRL's reviewer recommends that NSW DPE check with the authors of the guidelines to see if they are willing to correct their publication to accord with the statements made in the RTS.

In making this recommendation, WRL's reviewer highlights that he understands that geological structure was not incorporated into the the Hume Coal model. John Ross (2018) states in Appendix H, Pages 7 to 10 of the RTS that "*No geological structure was incorporated into the model, which is considered a reasonable approach for this Class 1 model*."

Therefore, given all the competing expert opinions by different practitioners on which of (now) two bureaucratic numbers should apply to the Hume Coal model, and given all the observations provided in Sections 3.1 to 3.9 of this response, WRL's reviewer re-affirms his view that the argument provided in the RTS over two numbers is a distraction.

Therefore, on this topic of model confidence, WRL's reviewer recommends that NSW DPE obtain expert opinion directly from the authors of the guidelines (Barnett et al., 2012) and the IESC on whether they believe:

1. That uncertainty in forward prediction is increased by the absence of model calibration to large hydraulic gradient and flow observation data (as they suggest it is);
2. The following draft advice by the IESC on uncertainty assessment (Middlemiss and Peeters, 2018) is reasonable (in the opinion of WRL's reviewer it is):

   > *"…geological model uncertainties become crucial in situations where groundwater models that are history-matched to head and discharge data for the historical*

pumping or climate record are then used for extrapolation beyond that conditional calibration base. In such 'out of range' simulations, the geological structure uncertainty may often be the dominant source, and thus alternative hydrogeological conceptualisations should form part of the uncertainty assessment."

3. The transient calibration is well supported by field data and what work might be undertaken to yield a better transient calibration; and
4. Any of the Monte-Carlo simulations reported in the RTS reproduce the structure and properties that could reasonably deduced from a detailed hydrogeological model of the geological and lithological logs and field and laboratory testing reported by Hume Coal using the method described in Section 3.2 of this review (actually undertaking such modelling work takes some time).

4.6 Data Sharing and Ownership of Hydrogeological Data

Finally, it remains unclear why Hume Coal has not fully documented or presented their 3D geological model and supporting data of the state owned Hawkesbury Sandstone Aquifer resource in the public domain. Similiarly, it is of concern to WRL's reviewer why NSW Government Department of Industry Water who has a mandate under NSW law to manage the groundwater resource are stated by Coal Free Southern Highlands Inc not to have a copy of the Hume Coal field data, the geological model or the numerical model.

WRL's reviewer understands that NSW owns all resources beneath the ground. Therefore, while Hume Coal has a licence to characterise the geological, mineral and water resources for the benefit of NSW in return for first right of access to share in profits from the resource (should the NSW choose to develop the resource), he does not understand the ethics of why the data on water resources and geology (i.e. excluding the coal resource) should be secret to Hume Coal. WRL's reviewer requests that the assessment report for the Hume Coal Project clarify NSW law and the conditions of the Hume Coal mining lease with respect to what field data, geological models, and numerical hydrogeological models remain secret to Hume Coal, NSW DPE and NSW DI Water.