Automated 3D physical property model builder for sparse constraints

NICHOLAS C. WILLIAMS

Mineral Deposit Research Unit
The University of British Columbia
6339 Stores Road
Vancouver, BC, V6T 1Z4
Canada

Geoscience Australia
GPO Box 378
Canberra, ACT, 2601
Australia

INTRODUCTION

Inversion of geophysical data seeks to extract a model, or suite of models, representing the subsurface physical property contrasts that can explain an observed geophysical dataset. As the observed geophysical data are a finite sampling of a continuous response, any recovered property distribution is only one of an infinite number of possible distributions that could explain the observed data. The most desirable solutions are those that can explain the observed geophysical data yet also reproduce known geological features; a goal that can only be achieved by including any available geological information into the inversions.

There are two approaches that can be used to include these geological constraints, based on the type of geological information available and the geological problem being addressed. A hypothesis testing approach supplies a full 3D model of geological interpretations and inferences to the inversion to test the hypothesis that those interpretations are consistent with the geophysical data (McGaughey, 2007; McInerney et al., 2007; Oldenburg and Pratt, 2007). Typically a qualitative assessment of the result is made based on how far the recovered inversion model deviated from the supplied interpretations in order to explain the observed geophysical data. However, in portions of the model that have low sensitivity to the geophysical data and no geological controls it may be possible to recover a property distribution that explains the observed geophysical data and is consistent with a flawed geological model. There will be no indication that the result is incorrect. This is particularly troublesome for gravity and magnetic data which are inherently non-unique due to the behaviour of potential fields. The gravity and magnetic responses decay with distance-squared and distance-cubed, respectively, so the sensitivity of subsurface model cells to surface geophysical data decreases dramatically with depth in the model. The amount of geological information available, and therefore the reliability of any 3D models produced also decreases significantly with depth. In precisely the areas where more reliable constraint information is required, less reliable constraint information is provided. In addition, as geological interpretations are required prior to performing any inversions, a significant amount of geological knowledge must be available, and a significant time commitment is required before any inversion results are obtained.

An alternate approach is to supply all available raw geological information to the inversion to recover a prediction about the subsurface distribution of geological features that may be required to satisfy both the known geological constraints and the observed geophysical data. This sparse data approach is particularly suited to problems where geological information is limited, sparsely distributed, or concentrated within restricted areas such as known ore bodies or along the surface. In such areas it can be difficult or impossible to build 3D models that are reliable enough to be included in the hypothesis testing approach.

This approach also postpones the geological interpretation until after the inversions have been performed, reducing the lead time to recovering an inversion result and enabling the results of inversions to be used in decisions to acquire further data.

This paper describes a method for preparing the geological constraints required for this sparse data approach eliminating the need for interpreting geology in regions of a 3D model that have limited or no geological information on which to base the interpretations. Applying geological constraints in any geophysical inversion procedure requires solid knowledge of the physical properties of the rocks. The technique outlined here specifically makes that physical property knowledge the central link between the geological information available and the geological constraints to be applied. In regions where little geological information is available, accurate physical property information may be unavailable, but the ability to include rough physical property estimates allows creation of simple models of geological constraints. Routine acquisition of physical property measurements during ongoing work will be rewarded by more rigorous and more robust constraint models which will provide more reliable inversion results on which to base further data collection. The technique also seeks to: 1) reduce the number of software packages required to integrate a variety of spatial datasets into a single physical property model by reading directly from raw data files; 2) automate the process as much as possible; and 3) ensure that the model can be updated quickly and easily when additional data becomes available so as to enhance the subsurface predictions. A demonstration version of the program with a basic graphical user interface has been developed using the Matlab programming package (Mathworks Inc., Natick, Massachusetts).

The method for developing a model of sparse geological constraints is specifically targeted for use with the UBC–GIF GRAV3D and MAG3D gravity and magnetic inversion programs (Li and Oldenburg, 1996, 1998) which allow geological constraints to be enforced on a cell-by-cell basis within a discretised volume. As described in the UBC–GIF GRAV3D and MAG3D inversion program manuals (UBC–GIF, 2005a, b), geological constraints are assigned to each cell by defining five parameters:

- A reference physical property which provides the best estimate of the actual physical property of the cell.
- A smallness weight which provides an estimate of the reliability of the assigned reference physical property. The weight is a unitless value ≥ 1 with increasing values indicating higher confidence. The default value of unity indicates that the reference physical property is uncertain and should not be strongly enforced in the inversion.
- A lower physical property bound which provides an absolute limit on the minimum physical property that can be assigned to the cell.
- An upper physical property bound which provides an absolute limit on the maximum physical property that can be assigned to the cell.
- Smoothness weights controlling the variation of properties between each cell in each direction. These weights are currently not handled by the model building routine described here, and are assumed to have default values of unity.

The lower and upper bound are effectively the confidence intervals around the selected reference property. The property recovered by an inversion for a cell that contains abundant geological information can therefore be constrained by assigning a reference physical property with a high smallness weight and a tight bounds range. Cells with no geological

information should be assigned some default reference physical property (associated with a density contrast of zero), a smallness weight of unity, and wide bounds that reflect the full range of possible physical properties. The inversion will recover a physical property model with properties for each cell that lie between the defined bounds, and are as close as possible to the supplied reference physical properties, while still reproducing the observed geophysical data. If possible, the reference physical properties will be matched more closely in those cells that have the highest smallness weights.

METHOD OVERVIEW

The fundamental goal of this approach is to create a 3D physical property model, based only on sparsely distributed raw geological data and restricted to those areas where information is available, using a repeatable, automated process. This physical property model can be used on its own to provide an additional tool to aid direct interpretation of subsurface geological data, or can be supplied to the UBC–GIF inversion programs as a constraining reference model. The inversion constraint parameters including reference properties, smallness weights, and lower and upper property bounds, are all related so all are created at the same time using the same data. The discretisation of the model, including the sizes and positions of cells, is defined by a UBC–GIF format mesh file which is supplied to the model building routine.

There are two main styles of observation that can be utilised in building a physical property model from geological data: actual physical property measurements and observations of rock types or alteration styles. Actual physical property measurements are obviously the most directly related to building a physical property model, however they are rarely taken systematically. Observations of geology are far more common. Since many geological units and rocks types have distinctive physical properties, observations of rock types and alteration may be used as a proxy for actual property measurements. A key component of building a physical property model that is partially based on rock type observations is to link the geological observations to appropriate physical property information. This is done early in the model building process via the creation of a physical property database for the model.

Once the physical property database is created, the model building routine can load the various data files containing those observations and extract or calculate the 3D coordinates at which the observations occur. The data files that can be supplied to the model building program are listed in Table 1. If the observation is a geological observation of rock or alteration types it is converted to a physical property estimate based on the physical property database. All of the physical properties are then used to populate the relevant cells in the model by combining all of the most reliable property measurements or estimates in each cell and extracting a statistical estimate of the most likely physical property value. Any cell that contains no geological information has default properties applied.

Table 1. List of the data types handled by the model building application. The data types are applied in the order of priority listed, which corresponds to their relative reliability within the model.

Data type	File formats	Description	Requires physical property database	Priority
Physical property measurements on surface samples	Column-delimited text files	Measurements taken on outcrop rocks or hand samples for which 3D coordinates are available	N	1
Physical property measurements on drill core	Column-delimited text files	Measurements taken on drill core samples, or by drill hole property logging tools, whose position is reported as a down-hole depth	N	2
Geology observations on drill core	Column-delimited text files	Observations of rock types and/or alteration styles taken on drill core samples, or by drill hole property logging tools, whose position is reported as a down-hole depth	Y	3
Outcrop or surface geology maps	ESRI vector polygon shapefiles	A nontopological shape format of vector coordinates of polygons and attributes (Environmental Systems Research Institute, 1998) storing observations of rock types, geological unit names, or descriptions made on surface rocks	Y	4
Basement or solid geology maps	ESRI vector polygon shapefiles	A nontopological shape format of vector coordinates of polygons and attributes (Environmental Systems Research Institute, 1998) storing interpretations of rock types or geological unit names expected at some position in the subsurface	Y	5
3D models of geological units	UBC–GIF inversion model format text files	A 3D model of geological units bounded by well-defined contacts, stored as a set of lithology IDs for each cell in the model volume	Y	6
3D models of geological domains	UBC–GIF inversion model format text files	A 3D model of geological domains that may span multiple geological map units and have poorly-defined boundaries, stored as a set of domain IDs for each cell in the model volume	N	7

An optional final step is available to expand the number of cells for which physical properties are assigned, if desired, by extrapolating the properties outwards from the data-bearing cells into adjacent buffer cells, based on the assumption that the physical properties will be roughly similar in adjacent cells. This assumption may be valid in some situations but not others, depending on the complexity of the geology, so the distance of extrapolation and the methods for determining the properties in the buffer cells are options for the user.

CREATING A PHYSICAL PROPERTY DATABASE

If a user wishes to include geology maps, drilling logs, or a 3D geological model as constraints, a physical property database must be created to provide estimates of the physical

properties associated with the observed geological labels. The database is created automatically from the following four data sources as depicted in Figure 1:

1. Physical property measurements and geology observations for surface samples
2. Physical property measurements and geology observations from drilling
3. A translation table that matches geology identifiers used in drilling logs to geology identifiers used in maps. This is necessary because drilling geology logs commonly use abbreviations and sometimes cryptic letter or number codes whereas maps tend to be more formal products with formal geological unit names, descriptions or identifiers.
4. Manually-specified properties for any geological rock types that are poorly sampled or difficult to measure properties on.

When creating the database, actual 3D coordinates of the observations are not required, only the co-location of property and geology observations is needed. The database assumes that the geology labels capture all characteristics that may control the physical properties of the rock, including weathering, metamorphism and alteration or mineralisation. If the geology labels apply to an inferred protolith rather than the actual rock, then any occurrences of the actual protolith will be assigned properties that more accurately represent the modified rock. Physical property measurements and geology observations for surface samples will typically be available in a single database, spreadsheet or table which can be extracted easily.

Drilling information

Drilling observations and property measurements require more careful manipulation as they will typically be stored in multiple databases or tables and will commonly have been recorded at different times by different people. The drilling information required for building this physical property model will be available in four separate tables:

1. Table containing coordinates of the drill hole collars and lengths
2. Table of drill hole survey orientations with measurements of the drilling azimuth and inclination at intervals down each hole.
3. Table listing physical property measurements at intervals down each hole
4. Table listing geological observations at intervals down each hole

Only the last two tables are required for creating the physical property database as actual spatial locations are not needed, only co-location of the geology and property observations. The data may include duplicate entries or overlapping observation intervals (e.g. if a hole has been logged more than once) which need to be handled carefully. Duplicate entries are skipped. If overlapping intervals exist, where the end point of one interval lies below the starting point of another interval, all interval endpoints are taken and sorted to create a new set of intervals that do not overlap, and the observation or measurements associated with the shortest original observation interval present at each new interval position is taken for the new interval (Figure 2). The shortest interval is used because it is assumed to represent the most detailed observation.

1 Extract all physical property measurements

Surface sample measurements

GEOLOGYCODE	DENSITY(G/CM^3)
Aod	2.93
Afv	1.75
Afv	2.35
Abm/Czl	1.99

Drill hole measurements

HOLEID	FROM	TO	DENSITY(G/CM^3)
ADJD406	86.0	88.0	2.15
ADJD406	118.0	120.0	2.58
ADJD406	212.0	214.0	2.51
ADJD406	214.0	216.0	2.51

Drill hole geology logs

HOLEID	FROM	TO	GEOLOGYCODE
LRC885	100.0	102.0	MA
LRC885	102.0	106.0	UK-
LRC885	106.0	110.0	MA
LRC885	110.0	112.0	FT

2 Assign property measurements to related map labels

Translation table

MAP_LABEL	DRILLING/SAMPLE_CODES
Dacitic porphyritic volcanic rocks	FP%
Basalt	MB%, MA%, Abb, Abv, Abmgp
Basalt - Plagioclase Phyric	MB%, MA%, Abb, Abv, Abmgp
Black Shale	SH, SB, $SH, $SB

3 Calculate statistics of all measurements for each label

For reference property

Mean
Median
Geometric ($\log_{10}$) mean

For lower and upper bounds (only if sufficiently sampled)

Minimum and maximum
±1, 2, or 3 standard deviations from mean
±1, 2, or 3 standard deviations from geometric mean
Percentiles (5th-95th; 10th-90th; 20th-80th)

4 Add manually-defined properties where needed

Translation table

LABEL	LOWER_BOUND(G/CM^3)	REFERENCE(G/CM^3)	UPPER_BOUND(G/CM^3)
Aog	2.25	2.94	3.2
Cza	1.4	1.9	2.3
Czb	1.4	1.9	2.3
Czc	1.4	1.9	2.5

5 Completed physical property database

Physical property database

LABEL	#SAMPLES	MEDIAN	MEAN	GEOMEAN	MINIMUM	...
Basalt	8562	2.83	2.90	2.89	1.31	...
FG	20	2.64	2.69	2.68	-	...
Granitoid/Gneiss	44	2.79	2.77	2.76	2.38	...
SH	53	2.82	2.86	2.85	2.32	...

Figure 1. Schematic depiction of the data inputs and stages for preparing the physical property database to be used to assign physical property estimates to geological observations throughout the model. 1. The data files are basic text files. 2. The translation table links the geology codes for surface samples and drilling logs to the labels used in maps. The wildcard '%' symbol matches all codes that start with the preceding letter combination to the corresponding map label. 3. Statistics are calculated to extract estimates of the reference property and lower and upper bounds associated with all measurements for each geology label. Bounds limits are only calculated if there are sufficient (≥ 30) measurements. 4. Manually-defined properties can be assigned for poorly sampled geology labels, or if the measurements are deemed inadequate for particular labels. 5. The finished physical property database provides a statistical summary of the physical properties associated with each geology label. Only a selection of the available statistics is shown. Note that the no minimum measurement is assigned to label FG because it has too few sample measurements.

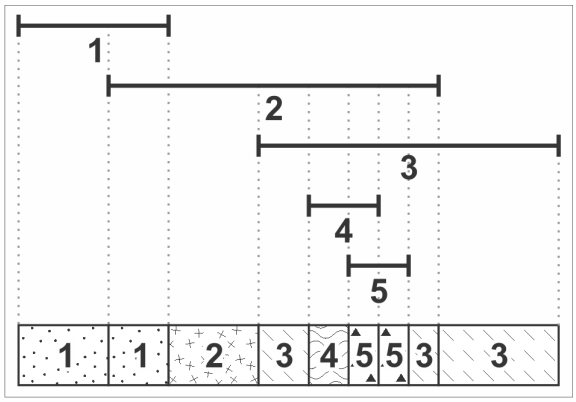

Figure 2. Schematic example showing how overlapping observation intervals in drill holes are handled. A new set of non-overlapping intervals are created from the end-points, and the observation associated with the shortest original interval at each position is extracted.

Since it cannot be assumed that the physical properties will be recorded at the same locations as the geology observations and will likely be recorded on different interval lengths, it is necessary to resample the observations on a regular user-defined interval. Any observations present at each resample point are extracted. All physical property observations are stored for later use in directly assigning property observations to model cells (Section 0). All points where both geology and property observations are present are extracted for use in creating the physical property database.

Translation table for geology labels

For the physical property database to apply to surface samples, drilling logs and geological maps, a facility is required to match any abbreviated geology codes and descriptors used for samples and geology logs with the more general or formal names commonly used in geological maps. This is accomplished via a text file translation table that specifies in one column the geology labels used in the maps, and in a second column any sample or drilling geology labels that correspond to those map labels. The implementation presented here allows wildcards to be used in the matching of labels to enable all labels starting with certain letter combinations to be easily grouped together into a single more general map label. All physical property observations from samples and drilling corresponding to each label (or wildcard) are combined and assigned to the identified map label.

Calculating property estimates

All the observed properties associated with each geology label identified in surface samples, drilling, or the translation table are compiled and a range of statistical measures are calculated on the properties associated with each label. All statistical measures for each label are stored for future use, so that the physical property database only needs to be rebuilt when new data are added, and can be reused with different selections of the statistical measures.

Estimates of the most likely property for each label are obtained using the median, the mean, and the geometric mean ($\log_{10}$) of all observations. Estimates of the variability in properties associated with each geology label are obtained using the minimum and maximum values, the mean ± 1, 2, or 3 standard deviations, the geometric ($\log_{10}$) mean ± 1, 2, or 3 standard deviations, and the 5-95th, 10-90th, and 20-80th percentile ranges. These variability estimates are used to ensure the variability inherent in the property estimates is carried through to the calculation of appropriate bounds (Section 0), so the user's choice will be

based on how tightly they want to constrain the properties. To ensure that reasonable variability estimates are obtained, they are only calculated if there is a sufficient number of samples; the number can be defined by the user, but a default value of 30 is used based on the Central Limit Theorem (Hoel, 1971).

Manually-specified properties

If physical property measurements are scarce, or do not cover the full suite of required geological labels, it is possible to manually assign the reference property and lower and bounds associated with any geological label. This is also useful if it is suspected that the measured properties for certain geological units are biased and unrepresentative. These manually-specified properties replace any automatically-calculated properties for those labels. Typically the manually-specified properties for each label, which are entered via a column-delimited text file, will be based on textbook property values, properties measured on similar rocks in other regions (although these measurements might be more usefully included in the drilling and sample data files and incorporated into the automatic property calculation), or by accessing some external physical property database such as the Mira Geoscience rock property database (http://www.mirageoscience.com/rpds: Parsons and McGaughey, 2007).

INPUT DATA TYPES

Once the physical properties for the geological labels are defined, the raw observation data can be imported and manipulated into the required format. Surface samples are again easy to include as the data will consist of a table of coordinates and physical property measurements that can be used directly.

Drilling information

As indicated in Section 0, the spatial position of the drill hole traces are defined by a file containing drill hole collar coordinates and hole lengths and a file of down-hole positional surveys indicating the azimuth and inclination of the hole at various depths. The physical property observations may be recorded on any interval length down the hole so overlapping intervals are reprocessed (Section 0; Figure 2) and the hole is resampled at a user defined interval with any physical property observation at each resample point extracted. The 3D coordinates for the resample points are calculated using the standard minimum curvature method for calculating the 3D coordinates of a sample point on a drill hole based on the positions and orientations of the adjacent drill hole survey points (Sawaryn and Thorogood, 2005). Minimum curvature assumes that adjacent survey points along the hole are connected by a circular arc centred at a point on the plane in which the drill hole trace lies (Sawaryn and Thorogood, 2005).The resulting data consists of a list of 3D coordinates and physical property measurements along each drill hole. The same procedure is performed for the drilling geology logs, with the observed geology code at each resample point extracted and the 3D coordinates calculated for each point.

Maps

Map data is commonly stored in polygon format ESRI shapefiles or can be translated into the shapefile format. Each polygon stored within the file contains coordinates for the geometry of the polygon, and attributes associated with the polygon. One of the attributes will be a geological label or descriptor which can be linked to the physical property database to

extract a property estimate for that polygon. The Matlab Mapping Toolbox (Mathworks Inc., Natick, Massachusetts) contains a built in utility to read the polygons and their attributes from shapefiles. The easting and northing coordinates for each column of cells in the model are defined by the supplied UBC–GIF mesh definition file; the horizontal centre of each cell is taken as the observation location for that cell. The vertical position of the observation location is taken from a UBC–GIF inversion model file that contains a topographic surface. The topographic surface can either represent true topography, or any defined surface that lies below topography, such as a basement horizon. A Matlab utility (interpshapefile: Kearney, 2006) interrogates the shapefile at the horizontal observation locations to identify which geology label is recorded for that location (if any) and a vertical coordinate is assigned based on the position of the topographic surface at that location. The properties associated with that geology label in the physical property database are then assigned to those coordinates.

Three dimensional geological and domain models

The purpose of this approach to building models from sparse constraints is to avoid the over interpretation required to build full 3D models. However, in some near-mine areas, enough subsurface geological information might be available to facilitate building a detailed geological model of units and contacts. This information can provide a useful framework for populating a physical property model, and the position of geological boundaries and contacts, if known well, can be powerful in defining regions of similar physical properties. Other model building methods generally use the full 3D geological model as the only, and therefore most reliable, constraint (Guillen et al., 2004; McGaughey, 2007). The method presented here only uses available 3D models as the least reliable constraint; all data and observations are applied over the top of the interpreted model. The geological model, created in a specialised 3D modelling package, must be available as a UBC–GIF model file format with an integer geological unit ID number stored in each cell. A separate text definition file defines the geological label associated with each of those ID numbers. The geological label is used to look up appropriate properties in the physical property database defined above.

Domain models are less reliable models of inferred geological characteristics for different regions of the model. In contrast to the geological models, the boundaries between domains may be poorly defined, and the domains may contain a range of lithologies and therefore a range of physical properties. Domains are particularly useful for assigning limits (bounds) to the physical properties for rocks in different regions of the model based on an understanding of the geological variability within those regions. One example is the transition from porous weathered material at surface to lithified basement rocks at depth. The exact depth of weathering may not be known, but an inference can be made that weathering does not extend below a certain depth. Any cells above that depth may have the low densities of weathered rocks or the higher densities associated with basement rocks, whereas all cells below that depth must have the higher densities associated with lithified basement rocks, and therefore a more restricted range of properties. The geophysical response will be used within the inversion to refine the thickness and extent of the weathered unit as needed. This situation is common enough that the model building routine includes an option to automatically build a weathering zone with a wider range of properties above a certain depth below the surface. This weathering domain will transgress any other defined domains or geological units. Domains are defined within a 3D model in the same way as geological units, with an integer domain ID number stored in each cell. However, the definition text file is instead a column-delimited file containing the geological unit IDs, and estimates of the lower bounds, reference property, upper bounds, and smallness weights (reliability) for each unit. Properties defined in

the geological and domain models are used as the default properties in the final model where no other information is available.

ASSIGNING DATA PROPERTIES TO THE MODEL

At this stage all of the input data has been reduced to 3D coordinates and either an observed property value or statistical estimates of the property value and bounds. If geological and/or domain models are supplied, their properties are applied as the default values for the model, with priority given to the geological model. If geological or domain models are not supplied, or are only defined for part of the model volume, user-defined default values are applied in the rest of the model. These will usually include relatively wide bounds and a smallness weight of unity which will ensure the reference model property in those unconstrained cells is not enforced as a constraint in the inversion.

The remaining data, physical property measurements and estimates from surface samples, drilling, and maps, are then used to assign a property to each cell. The model reference property is assigned based on the best available data in the cell as defined by the priorities in Table 1. Within each cell, all surface and drilling measurements and property estimates from drilling logs are combined using a user-defined statistic, either median, mean, geometric ($\log_{10}$) mean or a smallness-weighted average. The smallness-weighted average property value uses the smallness values assigned to each data type as weights so that the calculated reference model property is biased towards the most reliable data contained in the cell, as defined by the user. This is particularly useful where both property measurements and property estimates from drilling geology logs are included in the same cell. Although the geology logs provide useful information, they will not be as reliable as the raw measurements which should receive a stronger weighting. A reference property is assigned even if there is only one property measurement in a cell, as it is assumed to be the best estimator of the property of the cell in the absence of other information. The smallness weight assigned to the cell is taken from the maximum smallness weight associated with any of the data types contained within the cell as any additional data should only enhance the reliability of the cell's properties.

When assigning property bounds to the cell, the variability estimates (such as the minimum or maximum, standard deviations, and percentiles) calculated for properties associated with geological labels are also included with the reference property estimates from drilling logs and any measurements. The property values available for each cell will therefore be biased towards the more extreme values captured by the variability estimates, however this is acceptable given that the property estimates associated with geological labels are likely to be les reliable than actually measurements. Narrow, non-default property bounds are only assigned to a cell if the cell has been sufficiently sampled since they need to be accurate as they are an inflexible constraint in the inversion. The spatial distribution of surface sample and drilling data within the cell is determined by dividing the cell into blocks of a size that can be approximately represented by an individual measurement; the default is 10 m × 10 m × 10 m, but this can be adjusted if more or less variability in properties is expected. If the number of blocks containing samples is above some threshold (e.g. 75 %) then the cell has been sufficiently sampled and non-default bounds are applied. The bounds are calculated from the combined property measurements, property estimates, and variability estimates using a range of user-selectable statistics. The preferred method is to calculate the bounds using the Central Limit Theorem for estimating confidence intervals containing the population mean, which represents the mean of the infinite number of possible measurements within the cell, and therefore the average property of the cell:

$$\bar{x} \pm z_{(\alpha/2)} \frac{\sigma}{\sqrt{n}} \qquad (n \geq 30)$$
$$\bar{x} \pm t_{(\alpha/2)} \frac{\sigma}{\sqrt{n}} \qquad (n < 30) \qquad 1.$$

where $z_{(\alpha/2)}$ is the two-tailed Z-score for a confidence level of α and degrees of freedom n–1, $t_{(\alpha/2)}$ is the two-tailed Student's t-score, σ is the standard deviation associated with the available property measurements or estimates in the cell, and n is the number of property measurements or estimates in the cell (Borradaile, 2003). For $n \geq 30$, $t_{(\alpha/2)} \approx z_{(\alpha/2)}$ and approaches 1.96 and 2.58 for confidence levels (α) of 95 % and 99 %, respectively. This estimate of the likely range of reference properties for the cell provides a tighter range when more measurements are available or the standard deviation of the measurements is smaller.

If the cell contains no surface samples or drilling, the model values are extracted from the outcrop map or basement map, if either is present. If still no data is found for the cell, then the cell retains its default property values as defined in the geological model, the domain model, or as specified by the user. Once this has been performed for every cell in the model, the model is complete and can be saved as a UBC–GIF reference model, weight model and bounds model files ready to be included in an inversion.

EXTRAPOLATING PROPERTIES

The constraining model created thus far is based only on the data and is only enforced where data is available. In data-rich areas a significant number of the cells may be constrained. However, in data-poor environments, such as early exploration stages, few cells will have constraints. An option is provided to extrapolate the observed data outwards a short distance into surrounding cells. Geostatistics provides several methods for extrapolating the observed data to populate model cells, however the general case of populating a 3D physical property from varied and irregularly-distributed property and geological observations across multiple geological units with varied structural trends, is a special case requiring computationally complex universal kriging (Rendu, 1981). Stationarity of the property mean cannot be assumed and in complex geological environments, the implicit assumption of dependence of the properties of adjacent cells on each other may also be violated. Calculating semivariograms and kriging a 3D model of a million cells with even a few thousand property measurements or estimates (which may be available with only 10 well-sampled holes) is also computationally expensive. One simplifying approach might be to subdivide the model into subgroups of cells and apply random kriging to the properties of each cell in each subgroup assuming that samples are located randomly throughout each cell (Rendu, 1981). But it remains that in regions where observations are extremely sparse or even absent, kriging relies entirely on specific semivariogram models and drift functions that require geological knowledge that may not be available. Kriging also doesn't allow for subjective data quality estimates, such as those that are available from the user-defined smallness weights.

Given that the method proposed here is best suited to sparsely distributed observations, and to speed performance, a simpler statistical approach is currently employed based on distance weighting. The results are quicker to obtain and in data-poor regions will be at least as reliable as geostatistical solutions. In data-rich regions, slower geostatistical methods will likely provide more accurate and robust solutions, and this provides an avenue for future development. The currently implemented method calculates a buffer zone to

represent the zone of influence around each data cell. Buffers are applied in all directions around all cells that contain data from surface or drilling property measurements or drilling geology observations. For cells that only contain data from maps, buffers are only applied downwards and outwards as the maps are assumed to depict the rocks lying below the mapped surface. Buffers are not applied to domain or geological models as they have already populated a volume.

Each buffer around a cell extends either a user-specified distance or to the nearest data-bearing cell, whichever is closer in any direction. Each data-type can be associated with a different buffer distance to reflect how generally applicable the user believes the properties associated with each data type are. If no other data-bearing cells lie within the buffer, then the reference property of the data-bearing cell is applied as the best available estimate of the reference property for every cell in the buffer zone. Given that the confidence in that property estimate will decrease with distance from the data-bearing cell, the smallness weight assigned to each cell in the buffer is derived from the smallness weight associated with the data-bearing cell but weighted by inverse distance from the data-bearing cell using:

$$w_{s_i} = \left[\left(\frac{d_i}{d_{buffer}} \right)^2 \cdot \left(1 - w_{s_0} \right) \right] + w_{s_0} \qquad 2.$$

where w_{s_i} is the smallness weight assigned to cell i within the buffer, d_i is the distance from the central data-bearing cell to the centre of cell i, d_{buffer} is the distance to the edge of the buffer, and w_{s_0} is the smallness weight assigned to the central data cell. If the central data-bearing cell was sufficiently sampled to have narrow, non-default bounds assigned, then bounds are assigned to every cell in the buffer; however, the same inverse-distance weighting is used to widen the bounds range with increasing distance from the data-bearing cell until they reach default background values at the edge of the buffer:

$$b_i = \left[\left(\frac{d_i}{d_{buffer}} \right)^2 \cdot \left(b_{default} - b_0 \right) \right] + b_0 \qquad 3.$$

where b_i is the bounds value (either upper or lower) for cell i, $b_{default}$ is the default background bounds value, and b_0 is the bounds value associated with the data-bearing cell. So for the simple case where there are no adjacent data-bearing cells, the properties assigned to cells within the buffer are shown in Figure 3.

If multiple data-bearing cells are present within a buffer zone, then the buffers around each of the data-bearing cells will overlap. Any cells that lie between data-bearing cells must take properties that reflect the influence of each of the data-bearing cells. The contribution of each data-bearing cell to the properties of each buffer cell is determined by the distance from each data-bearing cell, and the smallness weight (or reliability) associated with each data cell's properties. This suggests a weighted average property with weights derived from both distances and smallness values. To be consistent with the inverse-distance squared weights used above, a reliability weight, R_{ij}, is calculated for each cell i in the buffer zone for the properties associated with each adjacent data-bearing cell j as defined by:

$$R_{ij} = \left(\frac{d_{buffer} - d_i}{d_{buffer}} \right)^2 \cdot \left(w_{s_j} - 1 \right) \qquad 4.$$

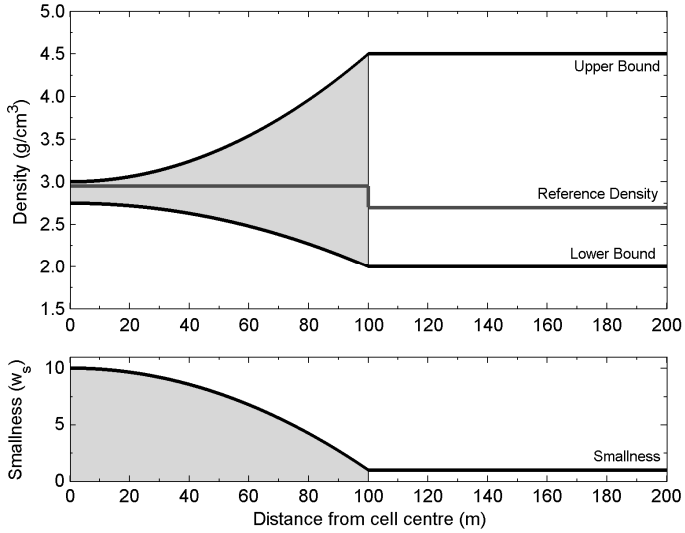

Figure 3. Unidirectional example of property values assigned within a 100-m-wide buffer surrounding a single data-bearing cell at 0 m. Within the buffer zone, shaded grey, all cells are assigned the reference property associated with the data-bearing cell, but have decreasing smallness weights and widening bounds as distance from the data increases and reliability of the property estimates decreases. Outside of the buffer zone (> 100 m) default values are used.

The $w_{s_j} - 1$ term adjusts the smallness weight of each data-bearing cell so that R_{ij} reaches a minimum of zero at the maximum buffer distance rather than the default smallness value of 1. Where the buffers of n data-bearing cells overlap, the weighted average property p extracted for each cell i is:

$$p_i = \frac{\sum_{j=1}^{n} R_{ij} \cdot p_j}{\sum_{j=1}^{n} R_{ij}} \qquad 5.$$

where p_j is the property of the jth data-bearing cell. The smallness weight assigned to each buffer cell is just the maximum R_{ij} for that cell plus 1 to ensure that $R_{ij} = 0$ corresponds to $w_s = 1$, the default value.

The remaining issue is to determine the distance of the buffer around each data-bearing cell to ensure that the buffer from one data-bearing cell does not extend beyond any adjacent data-bearing cells. The adjusted buffer distance, $\tilde{d}_{buffer_i}$, in the direction of $\mathbf{v_i}$ is calculated as a linear function of the angle α_{ij} formed between each buffer cell i, and each data-bearing cell j as shown in Figure 4 using:

$$\tilde{d}_{buffer_i} = d_{buffer} - \left(d_{buffer} - d_j\right) \cdot \cos\alpha_{ij}, \qquad \alpha_{ij} \leq 180° \qquad 6.$$

where:

$$\cos\alpha_{ij} = \frac{\mathbf{v_i}}{d_i} \cdot \frac{\mathbf{v_j}}{d_j} \qquad 7.$$

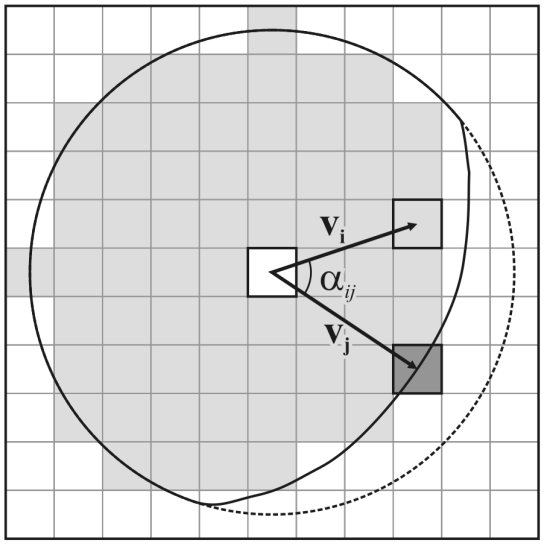

Figure 4. Two-dimensional example of how the size of the buffer around a data-bearing cell (centre cell) is reduced when other data-bearing cells lie within the buffer zone. The cosine of angle α_{ij}, between each buffer cell *i* (light grey shading), and each data cell *j* (dark grey shading), is calculated using Equation 7. A new buffer distance, $\tilde{d}_{buffer_i}$ (solid outline; compare to the user-defined buffer distance shown by the dashed line), is calculated using Equation 6. Potential buffer cells that have an angle $\alpha_{ij} \leq 180^{\circ}$ and $d_i \geq \tilde{d}_{buffer_i}$ are excised from the buffer zone. In practice the calculation is performed in three dimensions and works regardless how many data-bearing cells are present within the user-defined buffer zone and how complex their distribution may be, and ensures that the region of influence of a data-bearing cell never extends beyond another observation.

The adjusted buffer distances are calculated iteratively for each buffer cell–data cell pair within each buffer zone and the minimum buffer distance in every direction is retained. This ensures that the buffer zone, and therefore the central data-bearing cell's influence, never extends beyond adjacent data cells and the weighted properties calculated using Equation 5 only include those properties from the nearest data-bearing cells. Where many data-bearing cells lie with a buffer zone, the shape of the buffer can become very complex as demonstrated in Figure 5. In the extreme case where the central data-bearing cell is surrounded on all sides by other data-bearing cells, the adjusted buffer distance around that cell becomes less than the distance between cells so the buffer contains no cells and no extrapolation occurs. Properties are only extrapolated where no other data is available, and the extrapolation only extends a relatively short distance outwards with sharply decreasing certainties (represented by the width of the assigned bounds, and the smallness weights applied) as distance from the data increases (Figure 6). For these reasons this method of populating the model gains some intelligence over standard geostatistical techniques.

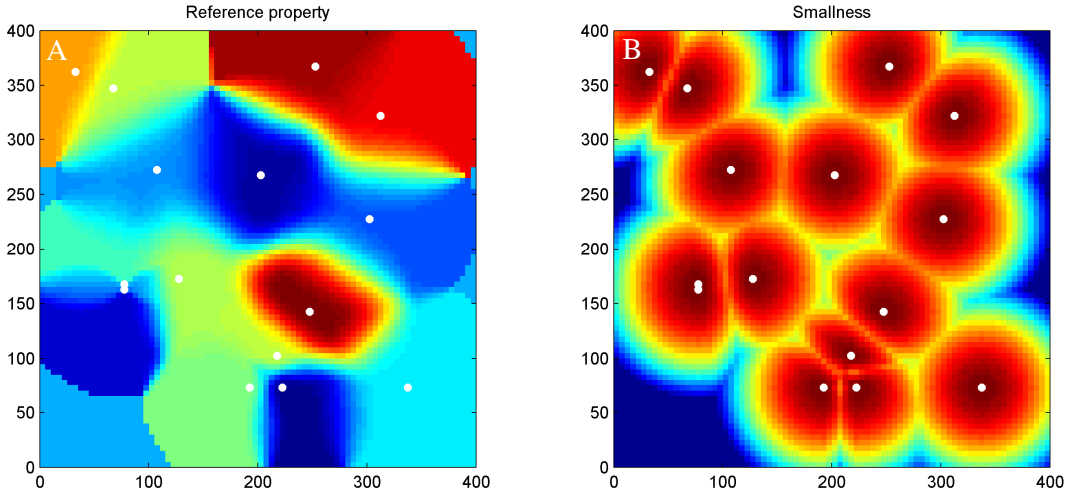

Figure 5. Example of buffers surrounding 15 random data points in a slice through a 3D model. The slice includes 80 × 5-m-wide cells in each direction. The data-bearing cells have randomly defined properties and their locations are marked by white dots. The default buffer width has been set at 100 m. A. Reference properties. Where a buffer contains no additional data cells, it has the default width and the same reference property defined throughout. Where two or more data-bearing cells lie within the proposed buffer, the buffer distance is reduced to the distance to the nearest data-bearing cell, and the reference properties of intervening buffer cells acquire a weighted average reference property from the adjacent cells. As a result, property gradients are observed between data-bearing cells. The steepness of the gradients depends on the separation between data cells. B. The smallness weights assigned to buffer cells decrease with distance from a maximum of 10 to a background level of 1 which is reached at the edge of the buffer. The width of the buffer in any direction is the minimum distance to either another data-bearing cell, or the user-defined buffer distance of 100 m.

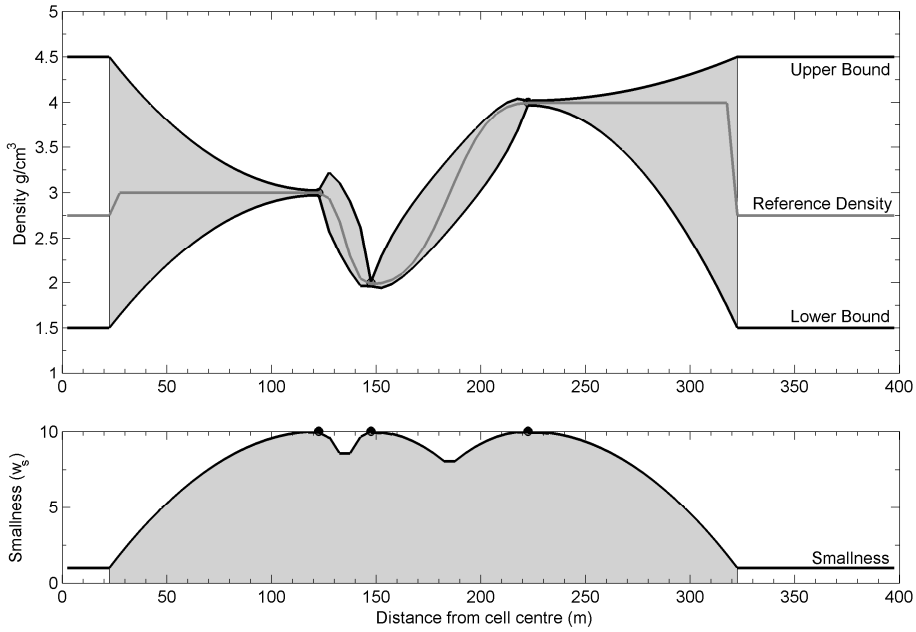

Figure 6. Unidirectional example of property values assigned within buffer zones surrounding three data-bearing cells marked with circles at 122.5 m, 147.5 m, and 222.5 m. At the data points, tight bounds and maximum smallness weights indicate a high confidence in the assigned properties. To the left and right ends of the profile the bounds widen to default values, and uniform reference properties are assigned. Between the three data-bearing cells the reference properties, bounds, and smallness weights are distance-squared-weighted averages.

EXAMPLE

A case study example was used to develop this sparse constraint model building program. Although confidentiality obligations prevent detailed analysis of the results in this report, some images can be used to illustrate the resulting model of sparse constraints. The case study seeks to use all available geological information surrounding an ore deposit to create a density model for constraining gravity inversions. The model volume measures 8.3 km east-west × 8.8 km north-south × 4 km vertically and has a minimum cell size of 25 m × 25 m × 10 m. The case study is particularly useful because it includes a large number of physical property measurements and geological observations in and around the ore deposit, but very few observations > 1 km away from the mine site. The geology is complex with multiply deformed and metamorphosed units, many of which are relatively thin. The irregular data spacing and complex geology represents a challenging problem for traditional kriging techniques, and highlights the benefits of the simpler and faster method presented here for populating the model.

The available data includes a Geological Survey of Western Australia (GSWA) 1:100,000 scale outcrop geology map shapefile, a company basement geology shapefile interpretation, a company's district-wide drilling database containing > 116,000 company density measurements and > 292,000 geology observations on > 34,000 drill holes, and density measurements on 30 variably weathered surface rocks from elsewhere in the district (N. Williams, unpub. data). The entire drilling database was included, even though many of the observations lie outside the volume interest, to improve the quality of the calculated density estimates in the physical property database, and to simplify the data input process; the model builder will extract only the data that it requires. No 3D geological model was used; however, a simple 3D two-domain model consisting of a regolith and Quaternary cover domain with a wide range of densities down to a depth of 100 m below surface, overlying an Archaean basement domain with more restrictive properties.

Basic model of surface constraints

Williams (in press) demonstrated that in areas where there is a strong physical property contrast between surface rocks and basement rocks, such as in deeply weathered terrain, providing a model constraining the physical properties of the surface rocks provides the most effective and easiest to implement constraint for enhancing the resolution of features at depth. The geophysical data are most sensitive to the physical properties of those surface rocks and so using surface constraints to improve the model at shallow levels can cause large changes at depth where the geophysical data has less of a control (Williams, in press). The simplest set of constraints that can be applied to surface rocks is to supply information from surface mapping with estimates of the physical properties associated with the map units. These types of constraints can be applied anywhere from prospect-scale delineation using company mapping, right through to regional scale targeting using government mapping. To demonstrate the approach for building a quick initial model of constraints using map data, much of the data in the case study area was ignored to simulate a greenfields exploration environment. A model was built using only the GSWA outcrop geology map shapefile (Figure 7A), the 30 surface sample density measurements, and estimates of physical properties based on literature values (Telford et al., 1990; Emerson et al., 2000) for 42 geology labels used in the outcrop geology map that had no available density measurements.

With these data inputs, the sparse constraint model builder created a reference model (Figure 7B), smallness weights, and bounds models, using 100-m-wide downward buffers to

extrapolate the map (Figure 7C), in less than 10 minutes. The resulting model contains constraint information for 12.2 % of the 805,291 model cells that lie below topography and clearly reproduces the mapped geology with islands of denser basement outcrops amongst dominantly low density weathered material. As described above, the buffers are calculated to ensure that the strength of the constraints, defined by the smallness weights and bounds range, decreases with depth away from the mapped surface.

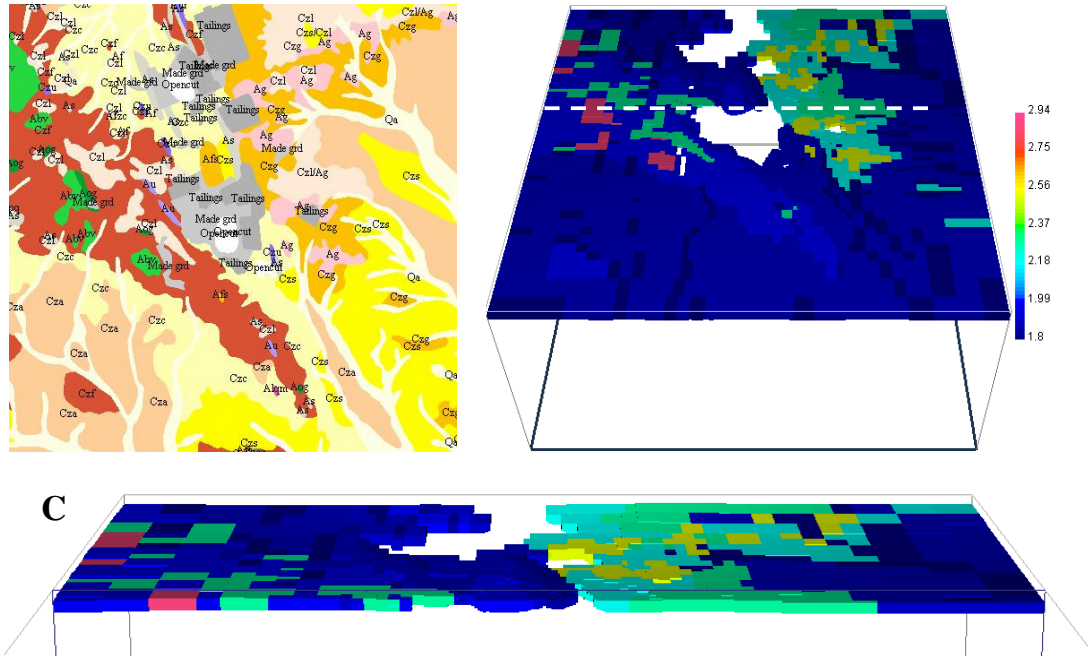

Figure 7. A. Portion of the GSWA outcrop geology map over the volume of interest. Measures 8.3 km east-west × 8.8 km north-south. The grey units represent mine infrastructure for which density estimates were not assigned. B. Resulting sparse constraint density reference model for the same region displayed in t/m^3, measuring 8.3 km east-west × 8.8 km north-south × 4 km vertically. Only those cells containing non-default values are displayed. The reference model clearly replicates the outcrop geology map with abundant low density weathered material with small patches of less-weathered basement outcrops. The dashed white line indicates the location of the slice depicted in C. C. Slice through the model showing the effect of extrapolating the map downwards using 100-m-wide buffers. At these shallow levels the mesh cells are 10 m tall, so a layer of 10 cells is constrained accounting for 12.2 % of the 805,291 model cells that lie below topography.

Complete model of sparse 3D constraints

When all the available geological information is included, much tighter constraints can be imposed over a much larger proportion of the model. All geology observations and density measurements were combined to create a database of physical property estimates for the region. Measurements were available for 1110 unique drilling or surface sample logging codes, but only 285 of those codes had sufficient numbers of measurements (≥ 30) to warrant calculation of variability estimates. A total of 49 basement geology map labels were not identified in the drilling logs and so were included via the translation table and automatically assigned properties based on analogous units in the drilling logs. Sufficient property measurements were available for variability estimates to be calculated on 44 of those. The same 42 manually-defined properties described above for the outcrop geology map were also included to complete the physical property database.

Although the entire database of drilling observations on > 34,000 holes was included in the physical property database calculation, only 5,400 of those drill holes lie within the

volume of interest. Only geology and property observations from this subset were used to actually populate the model. Likewise, none of the surface sample density measurements lay within the inversion volume and so were automatically excluded from the resulting model. Different smallness weights were assigned to each data type to reflect the varying reliability of each: 20 for drill hole density measurements (the most reliable properties), 15 for property estimates from drilling geology observations or the outcrop geology map (moderately reliable density estimates from reliable geology observations), 10 for basement the geology map (moderately reliable density estimates from an interpreted map), and 2 for cells in the two domains (purely estimated). These values are somewhat arbitrary and different values may be more appropriate in different areas and for different problems. Bounds were calculated for each data cell that had sufficient spatial sampling to accommodate 99% confidence intervals on the observed mean densities. In contrast to the simpler example outlined above, the 3D domain model is included in this model to give an estimate of the depth extent of weathering based on drilling. A more conservative 50-m-wide buffer is assigned to the outcrop map data to allow for changing densities with depth in the weathering profile. All other data are assigned 100-m-wide buffers.

A view of the resulting density reference model is shown in Figure 8A and the associated smallness weights indicating how strongly the reference model should be enforced in the inversion are shown in Figure 8B. The data types used to assign the properties to each cell are shown in Figure 9A for cells that contain actual observations, and in Figure 9B for cells that had properties assigned based on extrapolation within buffers. The contributions of all the data to the final reference model are outlined in Table 2. In this example, 20.7 % of the 805,291 model cells that lie below topography have constraints applied and the model took less than 50 minutes to build. In addition to providing a useful model for constraining inversions, the recovered reference model can even be used directly to make some inferences about the geology by synthesising a wide variety of geological observations into a common format, densities. The model has raised several testable questions about the distribution of alteration at depth, and provides a quick 3D volume representation of the complicated stratigraphy and structure in the subsurface.

Table 2. Breakdown of the number of cells constrained with information from each data type and whether the constraint came directly from the data or via buffer extrapolation. The model contains 920,856 cells, but only 805,291 of those lie below topography; cells above topography were ignored in these totals. Of most importance is that 3.5 % of the model can be constrained with data alone, and a total of 20.7 % of the model can be constrained by applying buffers around the observed data. All cells > 100 m from a data-bearing cell are only constrained as being either basements rocks, or possibly cover sequences.

Type of data	Model cells containing data		Model cells with properties from buffer extrapolation		Totals	
	Number	Percent	Number	Percent	Number	Percent
Surface sample density measurements [1]	0	0.0%	0	0.0%	0	0.0%
Drilling density measurements	4,127	0.5%	29,545	3.7%	33,672	4.2%
Drilling geology observations	6,462	0.8%	27,859	3.5%	34,321	4.3%
Outcrop map geology	7,008	0.9%	35,703	4.4%	42,711	5.3%
Basement map geology	10,477	1.3%	45,456	5.6%	55,933	6.9%
Total geologically-constrained	28,074	3.5%	138,563	17.2%	166,637	20.7%
Total geologically-unconstrained	777,217	96.5%	666,728	82.8%	638,654	79.3%

[1] Surface samples are included in the table because 30 samples were used to help calculate the physical properties database, but all lay outside the actual volume of interest.

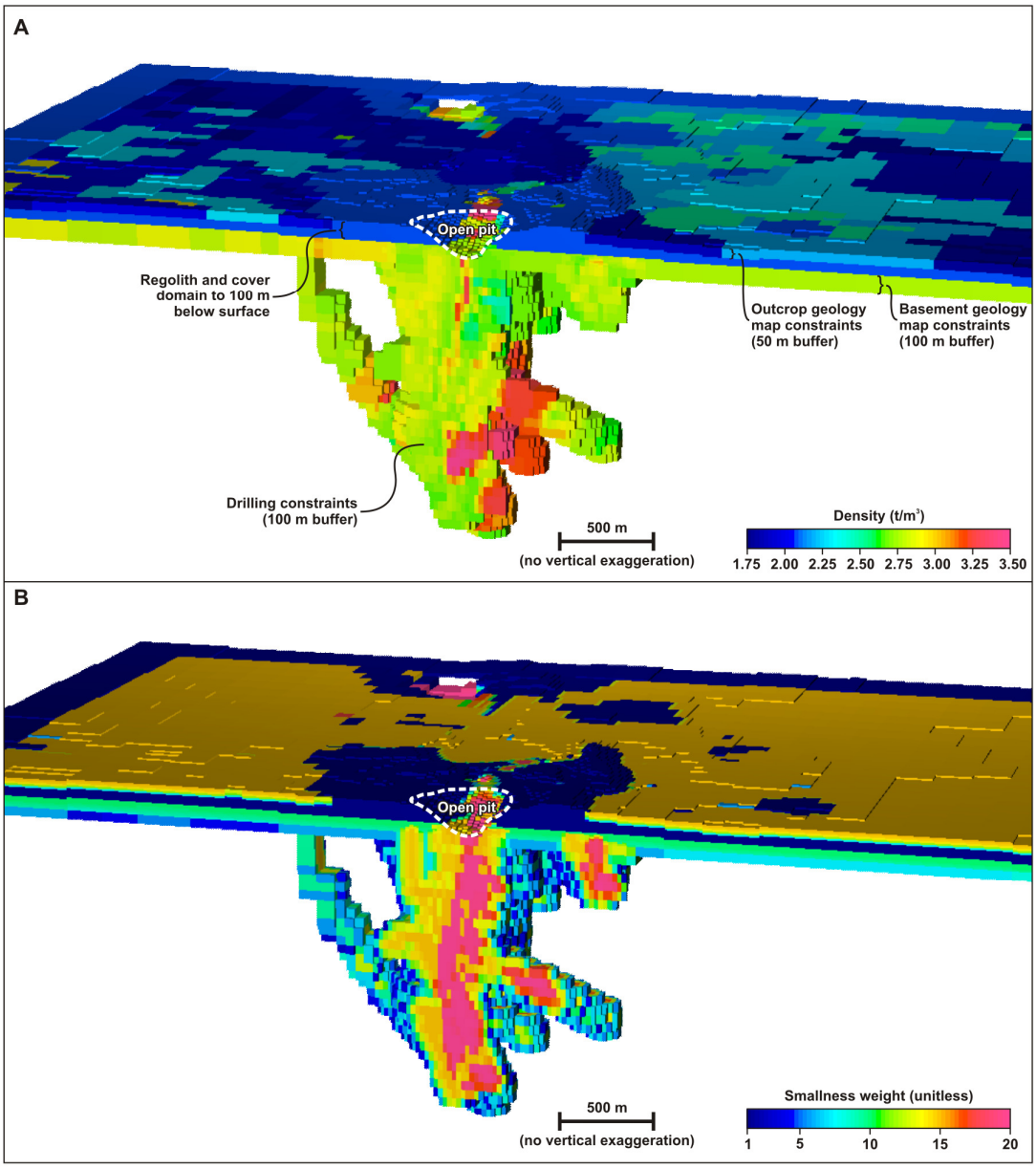

Figure 8. Cutaway perspective views of the final constraint model showing only those cells which were assigned non-default constraints. A. Reference density model showing the expected densities for each cell based on all available data and buffer-based extrapolation. Labels schematically indicate which data types were used in each area, but a more detailed identification of the data types is given in Figure 9. Geological structure is clearly present in the model and correlates extremely well with the company's wireframe surface interpretations (not shown and not included in the model building process). B. Smallness weights assigned within the model. These indicate the inferred reliability of the reference densities assigned to each cell. The most reliable data clearly lie at the surface and in the core of the main region of drilling, and reliability decreases outwards from the available data until it reaches default values (transparent cells).

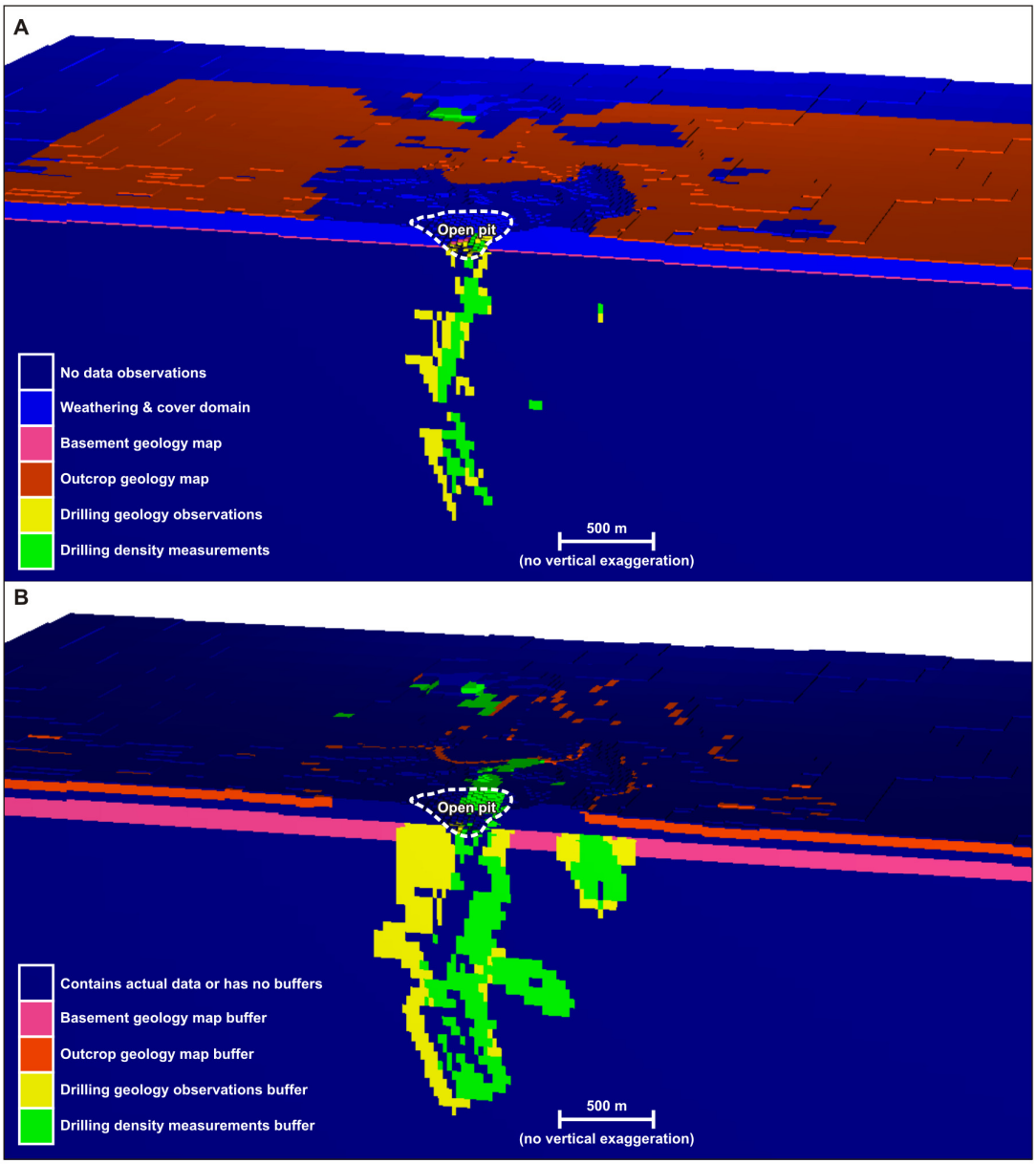

Figure 9. Cutaway perspective view at the same position as Figure 8, showing what the primary data source was for the properties assigned to each cell in the constraining model. For clarity, all cells are shown, not just those that had constraints applied. A. All cells that actually contained geological information, coloured by data type. B. All cells which were assigned constraints solely based on extrapolation of data within buffers. It is important to note that the buffers on the displayed slice are also affected by the 3D location of data in adjacent slices. The full constraining model shown in Figure 8 is the union of the contributions shown in A and B.

SUMMARY

The sparse constraint model builder described here provides a quick and efficient means of producing data-based constraining models for geophysical inversions, specifically targeted for use with the UBC–GIF inversion programs (UBC–GIF, 2005a, b). The procedure itself is primarily a data management routine to provide a systematic and repeatable way of combining geological observations and physical property measurements into a single, self-consistent model. Although many of the proposed steps can be achieved, or at least approximated, in other model building packages such as Gocad, they will typically require more time and effort to achieve. Physical property data is integral to applying the technique, and this may be perceived as a limiting factor in applying the method in some areas, but physical properties provide the critical link between geology and observed geophysical responses and physical property knowledge is a necessary component in any geophysical inversion method. By demonstrating an efficient link between physical property measurements and development of a constraining model for inversions, it should provide justification for acquiring more measurements in the field. But as demonstrated above, the simplest constraining model for any area can be achieved with a desktop study using a downloaded government geology map and physical property estimates sourced from literature values or online physical property databases. Any additional data acquired during a work program can be rapidly included as it is acquired to continually improve the working constraint model for a given area.

REFERENCES

Borradaile, G., 2003, Statistics of earth science data: Their distribution in space, time, and orientation: Berlin, Springer-Verlag, 351 p.

Emerson, D.W., Macnae, J., and Sattel, D., 2000, Physical properties of the regolith in the Lawlers area, Western Australia: Exploration Geophysics, v. 31, p. 229-235.

Environmental Systems Research Institute, 1998, ESRI Shapefile Technical Description: http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf, accessed 2 Nov 2007, Environmental Systems Research Institute.

Guillen, A., Courrioux, P., Calcagno, P., Lane, R., Lees, T., and McInerney, P., 2004, Constrained gravity 3D litho-inversion applied to Broken Hill: ASEG 17th Geophysical Conference & Exhibition, Sydney, 15-19 August.

Hoel, P.G., 1971, Introduction to mathematical statistics, 4th ed.: New York, Wiley, 422 p.

Kearney, K., 2006, interpshapefile, MATLAB Central File Exchange: http://www.mathworks.com/matlabcentral/fileexchange/, accessed 5 Nov 2007.

Li, Y., and Oldenburg, D.W., 1996, 3-D inversion of magnetic data: Geophysics, v. 61, p. 394-408.

Li, Y., and Oldenburg, D.W., 1998, 3-D inversion of gravity data: Geophysics, v. 63, p. 109-119.

McGaughey, J., 2007, Geological models, rock properties, and the 3D inversion of geophysical data, *in* Milkereit, B., ed., Proceedings of Exploration 07: Fifth Decennial International Conference on Mineral Exploration, p. 473-483.

McInerney, P., Goldberg, A., Calcagno, P., Courrioux, G., Guillen, A., and Seikel, R., 2007, Improved 3D geology modelling using an implicit function interpolator and forward modelling of potential field data, *in* Milkereit, B., ed., Proceedings of Exploration 07: Fifth Decennial International Conference on Mineral Exploration, p. 919-922.

Oldenburg, D.W., and Pratt, D.A., 2007, Geophysical inversion for mineral exploration: A decade of progress in theory and practice, *in* Milkereit, B., ed., Proceedings of Exploration 07: Fifth Decennial International Conference on Mineral Exploration, p. 61-95.

Parsons, S., and McGaughey, J., 2007, Rock property database system, *in* Milkereit, B., ed., Proceedings of Exploration 07: Fifth Decennial International Conference on Mineral Exploration, p. 933-938.

Rendu, J.-M., 1981, An introduction to geostatistical methods of mineral evaluation: Johannesburg, South African Institute of Mining and Metallurgy, 84 p.

Sawaryn, S.J., and Thorogood, J.L., 2005, A compendium of directional calculations based on the minimum curvature method: SPE Drilling & Completion, paper 84246, p. 24-36.

Telford, W.M., Geldart, L.P., and Sheriff, R.E., 1990, Applied Geophysics: New York, Cambridge University Press, 770 p.

UBC–GIF, 2005a, GRAV3D version 3.0: A program library for forward modelling and inversion of gravity data over 3D structures: The University of British Columbia–Geophysical Inversion Facility, 46 p.

UBC–GIF, 2005b, MAG3D version 4.0: A program library for forward modelling and inversion of magnetic data over 3D structures: The University of British Columbia–Geophysical Inversion Facility, 41 p.

Williams, N.C., in press, Applying geological constraints in UBC-GIF potential field inversions for greenfields or brownfields exploration, *in* Lane, R., and Fullagar, P., eds., Geologically-realistic inversions; Proceedings from the AESC conference workshop, July 2006 (tentative title): Geoscience Australia, Record.