Automated geological map deconstruction for 3D model construction

Mark Jessell[1], Vitaliy Ogarko[2,6], Mark Lindsay[1], Ranee Joshi[1], Agnieszka Piechocka[1,5], Lachlan Grose[3], Miguel de la Varga[4], Laurent Ailleres[3], Guillaume Pirot[1]

[1] Mineral Exploration Cooperative Research Centre, Centre for Exploration Targeting, School of Earth Sciences, The
5   University of Western Australia, Perth, Australia
[2] International Centre for Radio Astronomy Research, The University of Western Australia, Perth, Australia
[3] School of Earth, Atmosphere and Environment, Monash University
[4] Computational Geoscience and Reservoir Engineering, RWTH Aachen, Germany
[5] CSIRO, Mineral Resources – Discovery, ARRC, Kensington WA, Australia
10   [6] ARC Centre of Excellence for all Sky Astrophysics in 3 Dimensions (ASTRO 3D)

Correspondence to: Mark Jessell (mark.jessell@uwa.edu.au)

Abstract. We present two Python libraries (*map2loop* and *map2model*) which combine the observations available in digital geological maps with conceptual information, including assumptions regarding the subsurface extent of faults and plutons to

15   provide sufficient constraints to build a reasonable 3D geological model. At a regional scale, the best predictor for the 3D geology of the near-subsurface is often the information contained in a geological map. This remains true even after recognising that a map is also a model, with all the potential for hidden biases that this 'model' status implies. One challenge we face is the difficulty in reproducibly preparing input data for 3D geological models. The information stored in a map falls into three categories of geometric data: *positional data* such as the position of faults, intrusive and stratigraphic contacts; *gradient data*,

20   such as the dips of contacts or faults and *topological data*, such as the age relationships of faults and stratigraphic units, or their adjacency relationships. This work is being conducted within the Loop Consortium, in which algorithms are being developed that allow automatic deconstruction of a geological map to recover the necessary positional, gradient and topological data as inputs to different 3D geological modelling codes. This automation provides significant advantages: it reduces the time to first prototype models; it clearly separates the primary data from subsets produced from filtering via data reduction and

25   conceptual constraints; and provides a homogenous pathway to sensitivity analysis, uncertainty quantification and Value of Information studies. We use the example of the re-folded and faulted Hamersley Basin in Western Australia to demonstrate a complete workflow from data extraction to 3D modelling using two different Open Source 3D modelling engines: *GemPy* and *LoopStructural*.

1 Introduction

30 The 3D description and quantification of geometries displayed by deformed rocks has a long history (Sopwith, 1834; Argand, 1911; Ramsay 1967; Ragan 1968), however given the technologies available at the time, these were typically manual calculations extracted from photos or sketches. It has also long been recognised that a geological map and its legend provide more than just the distribution of lithological units but is a compendium of many different types of information (Varnes, 1974). Burns (1988) pioneered the analysis of maps in terms of the spatial and temporal relationships stored within, and Harrap (2001)

35 defined a legend language with aim of replacing list-style legends with diagrams that more clearly depict the historical, narrative aspect of geological mapping. Extracting information from digital GIS maps was pioneered in the context of mineral prospectivity (Bonham Carter, 1994), and more recently to validate the maps and analyse specific structures such as stratigraphic contacts and faults (Rauch et al., 2019; Kelka et al., 2020). 3D modelling packages often have basic data ingestion schemes that can import GIS data, for example the open source package *gemsis* (https://github.com/cgre-aachen/gemgis) is an

40 example of a system to speed up ingestion of data into the *gempy* 3D modelling platform, which assumes that the data is already in the fundamentally correct format (e.g. contact data has already been parsed to determine the base of the unit).

3D modelling workflows are slow, for the most part irreproducible, and the tracking of provenance of information leading to modelling choices is effectively impossible. In this study we present the first attempts at improving that part of the 3D

45 modelling workflow related to the transformation from map data to first model, which is one of the most time-consuming parts (hours to days) of the pre-model-building process. This study is aimed at hard-rock regional modelling scenarios which are generally data-poor compared to mines and sedimentary basins, and is part of the Loop project, a OneGeology consortium to build a new Open Source framework for 3D geological modelling (Ailleres et al., 2018; http://Loop3D.org). The aim of the libraries described here is to provide Loop and other 3D modelling systems with a unified method for accessing legacy digital

50 geological data, either from local files or online data servers, and extract the maximum geological information available for use as constraints on the 3D modelling process, as well as other related studies.

Commonly, the best predictor for the 3D geology of the subsurface is often the information contained in a geological map or if available logged well data. Unfortunately, away from basins and mines, drill-holes are often too shallow to provide constraints at the regional scale, and also often lack stratigraphic information. The information contained in a map falls into

55 three categories of geometric data: *positional data* such as the position of faults, intrusive and stratigraphic contacts; *gradient data*, such as the dips of contacts or faults. In a 3D workflow; spatial and temporal *topological data*, such as the age relationships between faults and stratigraphic units, we combine all of these direct observations with conceptual information, based on our understanding of the tectonic history of the region, including assumptions regarding the subsurface geometry of faults and plutons, to provide sufficient constraints to build a 3D geological model. Often, these conceptual assumptions are

60 communicated via geological cross-sections supplied with the map, however these are typically based on limited or no additional data and strongly rely on subjective and qualitative interpretations, although they can now routinely be validated using regional geophysical datasets such as gravity and magnetics.

In this study we attempt to reproduce manual workflows and structural decision-making process by developing a suite of algorithms that allow us to automatically deconstruct a geological map to recover the necessary positional, topological and

65 gradient data as inputs to different 3D geological modelling codes. Some of the code simply reproduces the 3D modelling packages' abilities to import different datasets, however much of it is dedicated to extracting additional information not previously available The codes described here retrieve information from GIS layers or online servers, clean and decimate the data if needed, and then go through a series of data analyses to extract information from these inputs, including: the local stratigraphy, the geometries of the basal contacts of units, and faults, estimates of local offsets along faults, and estimates of

70 local formation thickness. Once these and other information have been extracted they are output as standard formats so that the target 3D modelling systems can use them as is. One might want to automate these manual data manipulations for many

reasons including considerations of speed; reproducibility; and separation of data, concepts, and interpretations. Although the primary aim of this study was to provide information for 3D modelling workflows, some of the outputs may be useful for 2D analyses.

75   Starting from standard Geological Survey of Western Australia (GSWA) map products, and by extracting primary (e.g. stratigraphic contact location) and secondary (e.g. local formation thickness) geometric information, as well as fault and stratigraphic topological relationships, we are able to export a complete input file for two Open Source geomodelling packages (*GemPy* de la Varga et al., 2016; *LoopStructural*, Grose et al. this volume). In principle this workflow could be extended to work with other implicit modelling platforms such as EarthVision (Mayoraz et al., 1992), Gocad-SKUA (Mallet, 2004) and

80   Leapfrog (Cowan et al., 2003).), although the generated input dataset may contain data that are not considered in the modelling workflow proposed by some of the packages. The idea of extracting information to feed 3D modelling algorithms directly from other data sources such as satellite data has been previously demonstrated by Caumon et al. (2013) and Wellmann et al. (2019). A parallel study building libraries for automating information extraction from drill hole data is presented by Joshi et al. (this issue), so will not be discussed further here. Similarly although geological cross-sections can be handled by similar

85   methods to those that are described here, for simplicities sake we will not discuss them here.

For clarity, we refer to 'inputs' as the inputs to *map2loop* library and 'augmented data' as the products of *map2loop*. The augmented data in turn form the inputs to the target 3D geological modelling engines. All temporary inputs and outputs from the related *map2model* library are wrapped within the *map2loop* library.

The different commercial and Open Source modelling packages we have targeted use overlapping source of information but

90   distinct data formats in order to perform their modelling (Table 1). Some of the augmented data produced by the library are not (yet) explicitly required by any of the packages, but are useful datasets for contextual regional analysis and can provide some guidance for studies un-related to 3D modelling. Using a series of scripts contained with the Python *map2loop* library (https://github.com/Loop3D/map2loop) and the incorporated *map2model* library (https://github.com/Loop3D/map2model_cpp) we are able to convert the digital geological data into augmented data:

95   stratigraphically-coded 3D locations of structural dips of beds; fault and fold traces; and the stratigraphic base of each unit. A partner project led by the Geological Survey of Canada is developing a Knowledge Manager to support higher level information as a geoscience ontology to provide conceptual frameworks for modelling, aggregated petrophysical data and other basic knowledge of relevance to 3D modelling workflows (Brodaric et al., 2009; Ma and Fox, 2013).

In addition to the *map2model* library, *map2loop* depends on a number of external Open Source libraries, and in particular

100   draws heavily on *Geopandas* (to manage vector geospatial data; https://geopandas.org/), *Rasterio* (to manage raster geospatial data; https://github.com/mapbox/rasterio), *Networkx* (to manage network graphs; https://github.com/networkx/networkx) and *Shapely* (to manage 2D computational geometry; https://github.com/Toblerity/Shapely), and these libraries in turn depend on numerous other Open Source libraries.

2 Input Data

The *map2loop* library uses the *Geopandas* library to load data from several persistent formats (ESRI shapefiles, MapInfo tab files, JavaScript Object Notation (JSON) format files) and can also load data from Web Feature Services (WFS). Geospatial data can be in any standard coordinate reference system (assuming a European Petroleum Survey Group (EPSG) code is supplied, http://epsg.io).

110   In the example we present here, we use the 2016 1:500 000 Interpreted Bedrock Geology map of Western Australia and the WAROX outcrop database (GSWA, 2016) as sources of the data needed to build a first-pass model of the region around the Rocklea Dome and Turner Syncline in the Hamersley Region of Western Australia (Fig 1). The area consists of upright

refolded folds of Archean and Proterozoic stratigraphy overlying an Archean basement cut by over 50 NW-SE trending faults that form a part of the Nanjilgardy Fault System.

115   In the paper below, the descriptions are deliberately generic as the *map2loop* library uses a configuration file that allows the user to define which fields in the GIS layers contain which information. A Jupyter notebook (http://jupyter.org) allows this configuration file to be generated from the input layers. There are currently six types of inputs required to supply the necessary information to build 3D models (Fig. 1). Once the sources of data are defined, an initial verification of the data is performed to assure that the different information needed to perform the calculations is present. The minimum input data required to run

120   *map2loop* is described in Appendix 1.

2.1 Chronostratigraphic POLYGON and MULTIPOLYGON layer

This vector layer describes the chronostratigraphic information that underlies the complete geological map. Although 3D geological models can be built from purely lithostratigraphic maps, the implicit modelling schemes targeted by *map2loop* assume some knowledge of the stratigraphy. Not all maps follow a chronostratigraphic logic, for example for a map legend of

125   C-B-A (in decreasing age, Fig. 2) a local area of the map may actually show up-sequence orderings of the type C-B-A-B-A-B, and in order for a 3D model to be built they would have to be recoded as C-B1-A1-B2-A2-B3-A3. Of course the repetition of the A-B may be due to deformation (folding of the sequence, or thrust repetition), however it often just represents a level of stratigraphic detail considered unimportant at the scale of the map, or a deliberate avoidance of implying knowledge about the local stratigraphy. The chronostratigraphic polygon layer may also contain information on the surficial geology, but for more

130   regional analysis this is either ignored by the *map2loop* library, or a map that provides interpreted bedrock geology can be used. A prototype system that accounts for thicker cover sequences is available, but not discussed further here. The layer may contain a mixture of single POLYGONS, MULTIPOLYGONS (sets of POLYGONS with the same non-spatial attributes), and or POLYGONS with holes (also stored as MULTIPOLYGONS, Fig. 3). Each POLYGON needs to contain:

   a) a list of the ordered closed-loop x,y locations of the defining nodes,

135   b) a stratigraphic code or name at a lower hierarchical level (such as formation, member), which we will refer to as 'units' (since the choice of stratigraphic resolution is up to the user, and on a map polygons will often have different levels of stratigraphic coding),

   c) one or more higher-level stratigraphic definitions (such as group, supergroup, supersuite, province), which we will refer to as 'groups'

140   d) one or more lithological descriptions that help to determine if the unit is volcanic, a sill or other types of intrusions or other types of sedimentary rocks.

   e) optionally, but importantly, the maximum and minimum estimated ages of the fine-scale.

In the case study presented here we use the 2016 1:500 000 Interpreted Bedrock Geology stratigraphic polygons of Western

145   Australia (GSWA, 2016).

2.2 Fault POLYLINE and MULTIPOLYLINE layer

This vector layer describes the location, orientation and displacement information on mapped faults or narrow shear-zones at the surface. The layer may consist of a mixture of POLYLINE (groups of POLYLINES with the same non-spatial attributes). MULTIPOLYLINES are subsequently disaggregated into distinct polylines by the *map2loop* library to allow fault length and

150   orientation analysis to be correctly performed.

Each polyline needs to contain:

   a) a list of the ordered open-loop of x,y locations of the defining nodes,

   b) a unique identifier so that the fault can be labelled in some way,

    c)   optionally the dip and dip direction (or strike) of the fault can be stored at its midpoint.

155

In the case study presented here we use the 2016 1:500 000 Interpreted Bedrock Linear Features layer of Western Australia (GSWA, 2016), filtered by *map2loop* to extract the faults.

2.3 Fold axial trace polyline layer

This vector layer describes the location and polarity (anticline vs syncline) information on mapped fold axial traces, defined

160   by the intersection of the fold axial surface and the surface of the Earth. The layer may consist of a mixture of POLYLINE and MULTIPOLYLINES (groups of POLYLINES with the same non-spatial attributes).

Each polyline needs to contain:

    a)   a list of the ordered open-loop of x,y locations of the defining nodes,

    b)   a unique identifier so that the fold axial trace can be labelled in some way,

165   c)   the polarity of the fold axial trace (syncline, synform, anticline or antiform).

In the case study presented here we use the 2016 1:500 000 Interpreted Bedrock Interpreted Bedrock Linear Features layer of Western Australia (GSWA, 2016) , filtered by *map2loop* to extract the fold axial traces.

2.4 Bedding orientation point layer

170   This vector layer describes the local orientation of bedding, and is often missing from map packages, but can be found in the separate databases, or original field notebooks. It could also be estimated by photo-interpretation and/or three-point analysis. The layer may consist of POINTS.

Each point needs to contain:

    a)   a single x,y location of the defining node,

175   b)   dip information,

    c)   dip direction, or strike information, which we will refer to as 'azimuth' to avoid confusion,

    d)   the polarity of the bedding (upright or overturned).

In the case study presented here we use the 2016 WAROX outcrop database (GSWA, 2016).

180 **2.5 Reference Stratigraphy**

Some countries have developed national-level stratigraphic databases (such as the Australian Stratigraphic Units Database, ASUD, Geoscience Australia and Australian Stratigraphic Commission, 2017; https://asud.ga.gov.au/) that allow access to detailed stratigraphic information at the formation-level and above. The max-min ages for individual polygons mentioned in Section 2.1 would typically be derived from such a database. This information is typically non-spatial, however assuming that

185   the mapped chronostratigraphic polygons share the same coding as the national database, we can use this to augment polygons with the relative ages of any units in the map, which in turn help to define the local stratigraphy in the map area. The map2loop library currently uses a condensed form of the ASUD database that defines neighbouring stratigraphic relationships as pairs (A overlies B) to refine the local stratigraphy (Fig. 1b).

190 **2.6 Digital terrain model**

This grid layer, usually derived from the SRTM (Shuttle Radar Topography Mission; Farr et al., 2007) or GDEM (Aster Global Digital Elevation Map; NASA/JPL, 2009) datasets, or a fusion of both, provides a uniform coverage of surface topography

measurements over most of the continents. The *map2loop* library uses the Geoscience Australia server for 90m coverage in Australia (Geoscience Australia, 2016), the 1km global coverage offered by the Pacific Islands Ocean Observing System

195 (https://pae-paha.pacioos.hawaii.edu/thredds/dem.html?dataset=srtm30plus_v11_land) or the topography.org server for 90m or 30m coverage outside Australia, although there are a number of such servers now available, and the data is directly downloaded for the region of interest during the processing workflow.

In the case study presented here (Fig. 1c) we use the 90m version served by Geoscience Australia (Geoscience Australia, 2016).

200

3 Augmented Data Outputs

3.1 Positional Outputs

The *map2loop* library combines the inputs described in Section 2 in different combinations to produce a series of minimum necessary outputs as *csv, geotif* and *gml* format files that can be used directly by the target 3D geological modelling systems,

205 or as sources of analysis for 2D studies. This section outlines the high-level logic of how the different inputs are combined to produce information needed by the target 3DGM systems. These outputs are grouped by type: positional outputs, which provide information on the location and shape of features; gradient outputs, which provide information on the orientation of features; and, topological outputs that provide information on the spatial and temporal relationships between features. By features we mean contacts between units, faults, fold axial traces and bedding measurements. The specific positional, gradient

210 and topological outputs are in most cases calculated by combinations of the positional, gradient and topological inputs, and so the ordering below does not in general reflect the order in which these augmented data are produced by the *map2loop* library, so reference is made to data calculated in later sections. Ordering the sections by order of calculation results is useful to get an understanding of the data flow (Fig. 4) but also produces a rather confusing back and forth in text form as some data is incrementally modified as the workflow progresses. Example pseudocode for some of the calculations is included in

215 Appendix 2.

This paper focuses on two libraries, *map2model* (C++), and *map2loop (*Python). *map2model* performs a spatial and temporal topological analysis of the geological map, and *map2loop* further refines this analysis by including information from non-map sources, such as stratigraphic databases, acts as a wrapper for *map2model* and performs all other calculations.

3.1.1 DTM

220 The online Digital Terrain Model (DTM) servers either provide the information at a fixed x,y spatial resolution, or allow the client to subsample the data. For regional geological models a high resolution topography model is usually not needed, so a 90m or even 1km DTM is often sufficient for our needs. The *map2loop* library imports a subset of the global or national DTM, which are usually provided using a WGS84 projection. This is then reprojected using the *Rasterio* library to a meter or other non-degree based projection system. This distance preserving coordinate system is appropriate for our suite of modelling

225 packages that produce Cartesian models where the x,y and z coordinates use the same length units. The reprojected transformed DTM are stored as a *geotif* format file.

3.1.2 Basal contacts

The *map2loop* library currently uses the convention that stratigraphic contacts are labelled by the overlying unit in the stratigraphy, so that the contacts represent the bases of units, which we will refer to as basal contacts. Basal and intrusive

230 contacts (for the moment ignoring sills) are calculated using the intersection of neighbouring Chronostratigraphic polygons (Section 2.1). At the moment sill-like intrusive contacts are ignored, as they do not follow either massive pluton-like geometries

or strict stratigraphic relationships, but are the current subject of further study. In order to determine the label of the resulting polyline, we analyse the stratigraphic relationship between the two polygons using the information from the local stratigraphy calculated by the map2model library (Section 3.3.1):

235   a) if the two units are both volcano-sedimentary units, we assign the unit name of the younger unit,

b) if one of the units is intrusive (not a sill) and the other has a volcano-sedimentary origin, we assign the intrusive unit name if the intrusion is younger than the volcano-sedimentary unit, or the volcano-sedimentary unit if the intrusion is older,

c) if both units are intrusive (not sills) we assign the contact name to the younger unit.

240   d) If one or both of the units is a sill, we ignore the contact completely.

The x,y coordinates come from the intersection polylines, and can be decimated by taking every n^{th} node, the z value comes from the DTM. Outputs from *map2loop* consist of (Fig. 5a):

a) a series of x,y,z points,

b) unique stratigraphic name for each polyline, and

245   c) for each point the polarity of the contact (relative direction of younging and dip direction, a value of 1 means they are in the same direction and hence the bedding is the right way up, for overturned beds the value is 0)

3.1.3 Fault position and dimensions

Processing of fault geometries consists essentially of extracting the x,y location of nodes, combining with the DTM to get z, and calculating the distance between fault tips to define overall fault dimensions. A lower fault length threshold can be applied

250   so that very short fault segments, which will have little impact on the model, can be ignored. A decimation factor that only stores every n^{th} node value can also be applied. If needed, prior to *map2loop* processing, we use *FracG* (Kelka et al., 2020) to recombine fault segments based on the coincidence of fault tip locations and similar orientations.

Outputs from *map2loop* consist of (Fig. 5b):

a) a series of x,y,z points

255   b) unique fault name for each polyline, and

c) for each fault polyline the dip, azimuth and length of the fault

3.1.4 Fold axial trace position and dimensions

Processing of fold axial trace geometries consists essentially of extracting the x,y location of nodes, combining with the DTM to get z. Fold polarity (anticline/syncline) is recovered and stored. A decimation factor that only stores every n^{th} node can be

260   applied. Outputs from *map2loop* consist of (Fig. 5c):

a) a series of x,y,z points

b) unique fold axial trace name for each polyline, and

c) for each fold axial trace polyline the polarity of the fold

3.1.5 Local unit thickness

265   The local apparent thickness of units is calculated by finding the intersection of a line normal to the local tangent of a stratigraphic contact and the next stratigraphic contact (Fig. 6). Based on the stratigraphic relationship there are three possibilities:

a) if the next contact is the stratigraphically adjacent and higher contact, the distance is calculated (T_a) and stored as a local apparent thickness measurement.

270   b) if the next contact is stratigraphically higher, but not the stratigraphically adjacent, the distance is calculated and stored as the minimum apparent thickness (T_m),

c) otherwise no calculation is made.

True actual and minimum thicknesses can then be calculated from the apparent actual and minimum thicknesses as:

275
$$T_t = T_a \sin(\theta)$$

where T_t is the true dip, T_a is the apparent dip and θ is the dip of the bedding (Fig. 6, Section 2.3.2).

As these calculations can potentially be made for each node of a stratigraphic contact, we often end up with multiple estimates per unit, for which we can calculate the aggregated information as follows:

280   a) if we have true actual thicknesses for a unit we do not calculate minimum thickness, but store the median and standard deviation of thicknesses, and use the median of the actual thicknesses to calculate the local normalised thickness for each calculated node.

b) if we only have minimum thicknesses we store the median and standard deviation of the minimum thicknesses, and use the median of the normalised thicknesses to calculate the local normalised thickness for each calculated node.

285   c) if we have neither actual nor minimum thicknesses, if needed we use the median of the medians of thicknesses of all units as a rough estimate of the thickness, and no normalisation is possible.

Outputs from *map2loop* consist of (Fig. 5d):

a) a series of x,y,z points

b) apparent, actual/minimum, normalised actual/minimum thicknesses for each node and error estimates where

290   appropriate

c) table of summary thicknesses for all units

3.1.6 Local fault displacement

We have implemented three distinct methods of estimating the displacement across faults, depending on data availability.

The most complete analysis of fault displacements is based on identifying equivalent stratigraphic contacts across a fault, and

295   measuring their apparent offset (Fig 6a, D_a). If we combine this with the local interpolated estimates of dip/azimuth for the whole map (Section 3.2.4), and we know the orientation of the slip vector, we can calculate the true fault offset (Fig. 6a, Fig. 5e). Unfortunately, slip vectors are often hard to measure in the field and rarely recorded in geological maps. Given this, we can make an arbitrary assumption that the slip vector is down-dip (F_t), and then calculate the displacement based on the dip of the bedding, and the dot product of the contact and fault trace normal as:

300
$$D_t = D_a \tan(\theta \; C_n \bullet F_n)$$

where D_t is the true displacement, D_a is the apparent displacement, C_n is the 2D contact normal, F_n is the 2D fault normal and θ is the dip of the bedding. Since these are local estimates, we can have multiple estimates along the same fault, in which case

305   even these poorly constrained displacement estimates are of interest, as the relative displacement pattern along the fault can still be determined. Where these displacement calculations can be made, we can also determine the local downthrown block by comparing the sense of displacement (dextral or sinistral) with the dip of the strata (Fig. 5h). Specifically, the downthrown direction is given by considering the cross product of the fault tangent, the contact normal and the sign of the relative offset as follows:

310
$$W = (F_t \times C_n)\text{sgn}(D_s)$$

Where W is the downthrow direction, F_t is the fault tangent, C_n is the contact normal and sgn(D_s) is the sign of the apparent displacement sense (positive is dextral). If W is negative, the downthrown direction is defined by the normal to the fault trace with a right hand rule, and if the result is positive, by the opposite direction.

315 The ability to match equivalent stratigraphic contacts across a fault depends on the type of geology, the scale of the project and the detail of the mapping. A second level of displacement estimates can be made by comparing the stratigraphic offset across the fault, so if we have a stratigraphy going from older to younger of C-B-A and a fault locally separates unit A and unit C, then we can assume the offset has to be at least the thickness of units B, so if we have estimates of unit thickness (see Section 3.1.5) then we can estimate minimum offset (Fig. 5f and Fig. 6b). If, for the same stratigraphy, the fault offsets the

320 same unit A-A, or stratigraphically adjacent units A-B, the conservative estimate of minimum displacement would be zero. Finally if we do not have unit thicknesses available, we can always simply record the stratigraphic offset in terms of number of units (Fig. 5g and Fig. 6b), so in the original example above, an A-C relationship across a fault can be recorded as a stratigraphic offset of 2. The last two methods are not currently used in the automated workflow to determine fault offset; however they do provide insights into which faults are the most important in a region.

3.2 Gradient outputs

325

In order to calculate, secondary information such as apparent fault throw (Section 3.1.6) and local unit thickness (Section 3.1.5), we calculate an interpolated bedding orientation field (Fig 7).

3.2.1 Bedding orientations

The orientation data produced by the *map2loop* library is derived from a combination of gradient and positional sources,

330 specifically the Bedding orientation point layer (x, y, dip, azimuth, polarity; Section 2.4), the DTM (z; Section 2.6) and the Chronostratigraphic polygon layer (unit; Section 2.1). A filter is applied to remove observations where the dip is zero, as our experience has shown that this usually reflects a measurement where the dip was unknown, rather than a true dip of zero. Of course if the zero was indeed a true measurement we could override this assumption. Optionally, the number of points can be decimated based on taking every n^{th} point from the layer. More sophisticated decimation procedures, such as those described

335 in Carmichael and Ailleres, 2016, for orientation data are the subject of current work. Internally the code uses a dip direction convention so if strike data are provided we convert these to dip direction before calculation.

Secondary gradient information can be assigned along all of the stratigraphic and intrusive contacts based on a series of simple assumptions:

a) the dip direction of all dips are assumed to be normal to the local tangent of the contact and are defined as zero at

340 North and positive clockwise.

b) the dip can either be uniformly defined, or for the case of stratigraphic contacts, based on interpolated dips (see Section 2.2.4).

c) the azimuth of intrusive contacts for dome- or saucer-shaped bodies can be arbitrarily be selected by choosing the polarity of the dips and the azimuth (domes have outward dips and inverse polarity, saucers have inward dips and

345 normal polarity).

3.2.2 Fold orientation

If fold axial traces are available, and in areas with otherwise sparse bedding information, it can be useful to seed the model with extra orientation information that guides the anticline-syncline geometries.

Outputs from *map2loop* consist of, for each fold (Fig. 7a):

350 a) x,y,z positions

b) a series of dip/azimuth pairs

c) stratigraphic unit for each position

3.2.3 Fault orientation

If fault orientation data is available, either as numeric dip/azimuth (e.g. dip value: 75, azimuth value: 055) or in text form (e.g.
355 dip value: 'Shallow, Medium, Steep, Vertical', azimuth value: Northeast) then this is recovered and stored, otherwise the orientation is calculated from the fault tips, and the dip is set to a fixed value or is allowed to vary randomly between upper and lower limits. In the absence of other supporting information the qualitative dip information assumes equally spaced dips between the shallowest and steepest term, and assumes that the shallowest term is not horizontal, so in the example above we would get 'Shallow'=22.5, 'Medium'=45, 'Steep'=67.5 and 'Vertical'=90.

360 Outputs from *map2loop* consist of, for each fault (Fig. 7b):

a) a dip/azimuth pair

3.2.4 Interpolated orientation field

It became apparent during the development of this library that obtaining an estimate of the dip from bedding everywhere in the map area was a necessary precursor to calculating important information such as unit thickness (Section 3.1.5), fault offset
365 (Section 3.1.6), as well as the dips of contacts at arbitrary locations. In an attempt to retain more geological control over the sub-surface geometries, De Kemp, (1998), used polynomial and hybrid B-spline interpolation techniques to extrapolate geological structure. All more recent 3D geological modelling packages involve generalised interpolants of one form or another (Wellmann and Caumon, 2018; and see Grose, this volume for a discussion of the strengths and weaknesses of the different interpolants). At the scale of the map, we observe that local bedding azimuth measurements are often relatively poor estimators
370 of the map-scale orientation field. This occurs because the point observations record second-order structures, such as parasitic folds. In order to avoid these issues we have instead chosen to use the primary orientation data only for dip magnitudes, for which we have no alternative, and use the azimuth of stratigraphic contacts as the best estimator of the regional azimuth field. To this end we calculate a regular dip field using a multiquadratic Radial Basis Function (RBF) of the primary orientation 3D direction cosines using the *scipy* library (Fig. 7c), and separately use an RBF to interpolate the 2D contact azimuth direction
375 cosines (l_c, m_c, Fig. 7d). Each set of orientations from structurally coherent 'super-groups' (see Section 2.4) are interpolated separately. For each super-group, we then combine these into a single direction cosine (l_o, m_o, n_o i.e. the direction cosines of the interpolated bedding orientations) taking the n_o value from the interpolated 3D direction cosines and the $l_c m_c$ terms from the 2D direction cosines and normalising so that the vector has a length of 1 (Fig. 7e). This gridded field is then available for the thickness and offset values as discussed above, but could conceivably be used with appropriate caution as additional
380 estimates of orientation in parts of the model where no direct observations are available, or for cross-validation with known values.

3.3 Topological outputs

The third class of modelling constraints derived by the *map2loop* algorithms use the spatial and temporal topology of the map layers. Specifically, it creates network diagrams showing the stratigraphic relationships between units in the region of interest
385 (Burns, 1988; Perrin and Rainaud, 2013; Thiele et al., 2016), network diagrams of the relationships between faults, and relationship tables showing whether a particular fault cuts a unit or group.

3.3.1 Local stratigraphy

The spatial and temporal relationships integrated into geological maps provide a key constraint for 3D geological modelling (Harrap, 2001; Perrin, 2013). At the scale of a map sheet, state/province or country stratigraphic legends are necessarily
390 simplified models of the complex range of stratigraphic relationships. Since our aim is to build a model for an arbitrary

geographic region, we need to be able to extract the local stratigraphic relationships rather than just relying on the high-level summaries. The *map2loop* library uses the *map2model* C++ library to extract local stratigraphic relationships from a geological map. *map2model* uses two of the layers sourced by *map2loop*: the chronostratigraphic polygon layer (Section 2.1), the fault polyline layer (Section 2.2).

395 Shared contacts between polygons defining units are labelled as either intrusive, stratigraphic or faulted based on the nature of the units either side of the contact, and the presence or absence of a spatially coincident fault polyline (Fig. 8). The logic is as follows:

  a) if a contact between units coincides spatially with a fault polyline, the contact is labelled as a fault contact

  b) if a contact is between one intrusive unit and a volcano-sedimentary unit, the contact is labelled intrusive if the
400 intrusive unit is younger than the other unit, or stratigraphic if it is older.

  c) if the contact is between two intrusive units, the contact is labelled as intrusive.

  d) Otherwise, the contact is labelled as stratigraphic….

The relative age of each unit is determined from the min/max ages supplied for each unit in the map, and if these are not available, or they have the same age, or age range, then no age relationship is assigned. The primary outputs from map2model
405 are a series of network graphs in Graph Meta Language formal (GML, Fig. 9a) that can be visualised by the free but not Open Source *yEd* package ((https://www.yworks.com/products/yed) or the Open Source *Gephi* package (https://gephi.org/). The stratigraphic relationship graph underpins the definition of local stratigraphy in the *map2loop* system.

As not all maps provide max/min age information, a refined stratigraphic ordering can be obtained by using a national or regional reference stratigraphic database (Section 2.5). Depending on the structure of the database, an age-sorted ordering of
410 all units in the database, or pairwise stratigraphic relationships, 'unit A overlies unit B', can be used to refine the ordering extracted from the map. The final local stratigraphy is built up progressively during the *map2loop* workflow as the different elements: basic relationships from the *map2model* library, followed by a refined version using a reference stratigraphic database to better characterise local age relationships. Even after these progressive refinements, ambiguities in relative age of units usually remain. At the moment *map2loop* arbitrarily choses one of the distinct stratigraphic orderings as the basis for its
415 calculations, but clearly this is an important source of uncertainty. A study is underway to better understand how stratigraphic uncertainty propagates into the resulting 3D geological uncertainty.

We can reduce the uncertainty in the stratigraphic ordering that comes from lack of information in the map as to relative ages, or ambiguous relative map age relationships, by considering one higher level of stratigraphy, which we will call 'groups' but could be any higher rank of classification. This reduces the uncertainty as typically the uncertainty in relative ages between
420 groups is smaller than the relative ages of any two units if we ignore their group relationships.

Since *map2loop* is primarily aimed at implicit modelling schemes, there is a considerable advantage in reducing the number of stratigraphic groups that have to be interpolated separately, since the more orientation data we have for a structurally coherent set of units the better the interpolation. To this end we use the *mplstereonet* Python library to compare each group's best-fit girdle to bedding orientation data so that if there are closely matched (a user-defined choice), they can be considered
425 to be part of the same 'super-group'.

The outputs of *map2loop* are a stratigraphic table defining a distinct ordering of units and groups, plus a table of which groups form super-groups to be co-interpolated.

3.3.2 Fault-fault relationships

The intersection relationships between pairs of faults are calculated by map2model by analysing which faults terminate on
430 another fault (Fig. 9b). This is assumed to represent an age relationship, with the fault that terminates assumed to be the older fault. The map2loop library converts this information into a table of binary relationships: Fault X truncates/has no relationship to Fault Y that are then compiled into a set of graphs of fault-fault relationships.

3.3.3 Fault-stratigraphy relationships

The intersection relationships between stratigraphic units and groups are calculated by the *map2model* library by analysing

435 which geological polygons have sections which are spatially coincident with faults. These are then converted by the *map2loop* library into two tables of the binary stratigraphic relationships unit/group A (rows in Fig. 9c) is cut by/is not cut by fault X (columns in Fig. 9c).

4. 3D Modelling of map2loop outputs

Once the augmented data types have been calculated by *map2loop* and *map2model*, a final validation of the data is performed

440 so that there are no 'orphan' data, for example orientation data for units that will not be modelled, and a unit in the stratigraphy for which we have no contacts or orientations. Although this can obviously happen in nature, current modelling systems struggle with this concept, so we need to ensure that the model will actually build by removing unresolvable data.. The outputs of *map2loop* and *map2model* described above provide all of the information required to build 3D geological models (Fig. 10), in *GemPy* (de la Varga et al, 2019) and *LoopStructural* (https://github.com/Loop3D/LoopStructural, Grose et al. this volume).

445 The ability to generate all necessary input data for a geological model from set of source layers in a matter of minutes demonstrates the potential for this approach to reduce the entry barrier for geologists who wish to make 3D models as part of their exploration or research programs. Of course the whole system relies on the quality of the input data, and when under cover, different and more geophysics-centric approaches need to be taken. The integration of geophysics into the workflow is being developed by the Loop consortium, but is beyond the scope of this paper.

450 **5. Discussion**

The choices made by the *map2loop* and *map2model* code attempt to reproduce the thinking of a geologist when manually building a 3D geological model from the same data. There are many small or large decisions and assumptions that are made when developing the model, and the discussion below highlights some of the areas where further work needs to be done to reproduce the manual workflow. In this paper we have used an example from Western Australia, however similar examples

455 for the Northern Territories, New South Wales, Vitoria, Queensland, Tasmania and South Australia can be run using the *map2loop* library). The example map and associated data used in this paper took just over 3 minutes to deconstruct with *map2loop* and a further 4-15 minutes to build with the three target modelling engines, running on a standard laptop computer.

5.1 Improvements to calculations

460 The aim of this study was to build an end-to-end workflow from raw map 'data' to a 3D model, which we hope to build upon by refining the different steps as discussed below.

5.1.1 Calculation of unit thickness

The calculation of local unit thickness (Section 3.1.5) depends on the local estimate of apparent unit thickness, which is reasonably robust, but also on the local estimate of the dip of the stratigraphy. This dip estimate comes from the application of

465 the *scipy* Radial Basis Function interpolation library, and in particular the multiquadratic radial basis function, which can be supplemented by a smoothing term. Other radial basis functions such as Gaussian and inverse are available, as well as other schemes such as Inverse Distance Weighting and co-kriging, which all offer multiple ways of estimating the local orientation field. We chose the multiquadratic RBF simply because our experience showed that, for the types of geology that we started

working on, it produced 'reasonable' results. It is likely that different geological scenarios may require optimised interpolation

470   schemes (Jessell et al, 2014) as there is no unique solution to this problem.

5.1.2 Calculation of fault offset

The calculation of local fault offset also relies on the interpolated dip field, so the same remarks regarding geologically appropriate interpolators stated in the previous section apply. If we compare the local displacements along a fault, then we also have to assume that the unit thickness is the same on both sides of the fault, but at least in general this can be tested directly.

475   In addition, to properly estimate fault displacement, we need to know the fault displacement vector. One solution, not yet implemented, would be to calculate the relative displacement of lines of intersection of the same dipping stratigraphic units either side of a fault, but this has not yet been implemented here.

5.1.3 Calculation of super-groups

The definition of super-groups for co-interpolation of bedding data is performed by comparing the orientation of best-fit

480   girdles. This has a number of flaws. Firstly disharmonic fold sequences may have the same orientation spread, but different wavelengths and thus should not be interpolated together. Secondly, if a particular group is undeformed, or lies on one limb of a fold, there may not be a well-developed girdle. A more robust analysis of fold structural information, which includes analysis of representative fold profiles, as described by Grose et al., 2019, would not only allow us to better identify coherent structural domains, but would also provide the information needed to use the more sophisticated modelling schemes described

485   in their work.

5.1.4 Choice of stratigraphic ordering

As described in section 3.3.1, the stratigraphic ordering of units is derived from a combination of local observations drawn from the geology polygon and fault polyline layers, and a regional or national reference stratigraphy. This process do not generally lead to unique stratigraphic orderings, and at present we simply take the first sorted result from a sometimes long

490   list of alternatives. A second unknown is the nature of the contact between different groups. We use the idea of super-groups to cluster structurally coherent domains, but we do not currently have a good solution to estimate the nature of discontinuities between structurally incoherent domains. The modelling systems we target allow for onlap and erode relationships, and Thiele et al. 2016 suggested the topological analysis of units to identify unique relationship characteristics between groups as a possible way forward, but this remains to be tested.

495   ### 5.1.5 Analysis of fault-fault and fault-unit topology

The assumption that a fault or unit that truncates against another fault represents an age relationship is reasonable, but exceptions obviously exists in reactivated faults and growth faults. At the present time if a cycle in fault age relationships is discovered: Fault A cuts Fault B; Fault B cuts Fault C; Fault C cuts Fault A, one of the age relationships is removed arbitrarily. A better approach may be to look at displacement, length or some other characteristic such as stratigraphic offset to make that

500   decision. A further measure may be the centrality of a fault, for which there are several methods (Freeman, 1977), for example related to how many other faults are truncated by a specific fault. These fault-fault and fault-unit age relationships could provide further constraints on the overall stratigraphic ordering of units, and of the structural history of a region that would be valuable inputs to time-aware modelling systems such as *LoopStructural*.

5.2 Limitations in resulting 3D models

505 Given the complexity of the task, and the limitations and somewhat arbitrary nature of some of the choices described above, it is perhaps surprising that we ever get a good 3D model out of the system. Conversely there are a number of other reasons why having deconstructed a map, we do not end up with a 3D model that meets our expectations or needs.

5.2.1 Insufficient data

All geological maps are models, as even in areas of 100% outcrop the map is the sum of hundreds of local observations and 510 interpretations, and in most areas the gaps in outcrop mean that the map can only provide a subset of the potential surface information. It may well be that the surface map does not possess enough information to constrain a 3D model. In many regions, the surface of the Earth is covered by soils or surficial deposits (colluvium, alluvium etc.) that prevent direct observation of the bedrock geology. In this case there is simply no map to deconstruct. As regional geophysical datasets became more widespread, interpreted maps of the top of bedrock started to be produced (such as the GSWA test case described here), 515 together with estimates of the geometry of the cover-bedrock interface (Ailleres and Betts, 1998). *map2loop* contains example code showing how these may be combined to replace the surface geology as inputs for modelling, but were not needed for the Hamersley test case. Even when surface geology maps are available, interpreted cross-sections are usually added to constrain the 3D geology, however even if they are constrained by geophysical data, by direct interpretation of seismic, or by gravity/magnetic validation for example, they are still usually less well-constrained than the surface data. Even when seismic 520 data is available, Bond et al., 2015 has shown that this prior experience is a significant source of bias for the interpreted section. Drill hole data are not currently incorporated into the workflow, however the work of Joshi et al., (this special issue) goes some way to providing that possibility. Geophysically unconstrained cross-sections drawn by geologists necessarily depend on two sources of information, the geology map, in which case in principal a future *map2loop* could provide the equivalent information, or by the geologists' prior experience, which is harder to codify, and represents a significant future challenge.

525 5.2.2 Poor quality data

The process of making a map, like any human endeavour, is subject to error, either as a result of the primary observation, or from the compilation of that information into map form. Some analysis of map logic can be made if the information in the input map or stratigraphy is incorrect, the fault cycles described in Section 3.3.2. If a 3D model fails to build using the deconstructed data, one may assume there are inconsistencies in the input data. The issue here is the modelling engine will 530 unlikely indicate which data is causing errors, so more robust map validator would be useful that can identify potential issues prior to 3D model input and provide guidance to correction. At present small mismatches between nodes in coincident polygons and polylines can be accommodated, however if one polygon or polyline has denser node spacing than a feature that is supposed to be coincident, we do not resolve these differences.

5.2.3 Incorrect deconstruction of the data

535 As discussed in section 4.1, *map2loop* makes a number of simplifications during the deconstruction process. Estimates of fault displacement and unit thickness could be checked for consistency along a contact or fault, which may improve the estimates fed to the modelling schemes.

5.2.4 Incomplete 3D modelling algorithms

The last reason that the outputs from *map2loop* do not always produce satisfying 3D geological models is that the modelling 540 systems themselves do not manage all types of geological scenarios well. The three modelling engines targeted here are all implicit schemes that work best in regions with a well-defined and gently deformed stratigraphy although *LoopStructural* can also handle poly-deformed terranes. Once overprinting of structures becomes more important, the implicit schemes need more

and more information (often provided as interpretations not directly supported by the original data) to reproduce the model conceived by the geologist. The conceptual model in the geologist's head is a major control on tuning the implicit model, and

545 codifying these concepts remains a major challenge for the future. To give just one example, the 3D geometry (and even the near-surface dips) of faults are often very poorly understood. In order to produce a 3D model a geologist often brings a preconceived notion: constant bed thickness, similar folding or chevron folding, extension related faults offsets with antithetic faults; compression related fault offset with low angle basal fault and associated folds with bedding thickness changes; transpressional and transtensional flower structures, which is then used to complete the model in an under constrained area.

550 All the regional scale "objects" (duplex, flower structures etc…) are basically fault networks that evolved with time, with complex slip histories. The LoopStructural library is specifically designed to tackle these sorts of evolutionary systems, however at present the challenge is that we have insufficient data to actually test it in real-world settings. If we could encode these concepts, then it would be easy to ask the automated system to compare model outcomes for the model "as if" it were an extensional listric tectonic environment vs a transtensional system.

555 One of the keys to improved modelling is to incorporate additional time constraints on the model. All three target modelling engines incorporate some concepts of time, such as stratigraphy-fault age relationships, and LoopStructural can handle superimposed fold and fault interference geometries if sufficient data is available (Grose et al., this volume). Finally, the choice of which data to put into the 3D model is by definition outside of the 'knowledge' of *map2loop*, as it can only process datasets it has been made aware of, however a broader data discovery algorithm that searched for all available data and then decided

560 on the basis of, for example, data density, relevance to question, volume of interest (Aitken et al., 2018) could be a way to avoid this currently biased process.

5.3 Future work

The enormous advantage of automating many of the somewhat arbitrary choices and calculations described in this paper is that alternatives can also be coded, and the sensitivity of the resulting 3D models to these choices can be analysed. Since the

565 process is automatic, the time taken to calculate 1000 models on a distributed computing system is the same as calculating one model, so very large model suites can be explored for very little additional time cost. This can build on existing capabilities: *GemPy* has its own advanced framework for analysing uncertainty (de la Varga et al., 2019). Work is currently underway to wrap the entire data extraction, 3D geological modelling and geophysical forward and inverse modelling workflow in Bayesian analysis framework, so that the distinct and cumulative effects of all modelling, uncertainty quantification and joint geological-

570 geophysical inversion decisions (Giraud et al., 2020) can be analysed in a homogeneous fashion.

In the immediate future *map2loop* and related codes need to manage a wider range of input datasets including drill holes and cross sections, and this work is underway. There is also a need to extract the maximum amount and range of information from sills and other igneous intrusions that do not follow simply stratigraphic or geometric rules. Perhaps the biggest challenge is the incorporation of conceptual constraints during the deconstruction workflow, as discussed in the previous section.

575 ### 6. Conclusions

map2loop automation provides significant advantages on manual 3D modelling workflows, since it:

- Significantly reduces the time to first prototype models;
- Allows reproducible modelling from raw data;
- Clearly separates the primary observations, interpretations, derived data and conceptual priors during the data
580 reduction steps and
- Provides a homogenous pathway to Sensitivity Analysis, Uncertainty Quantification, Multiscale Modelling and Value of Information studies.

7. Author Contribution

Mark Jessell was responsible for implementation of the computer code and supporting algorithms for the *map2loop* and wrote
the initial draft of the manuscript. Vitaliy Ogarko was responsible for implementation of the computer code and supporting
algorithms for the *map2model* code. Mark Lindsay tested the code on different datasets and contributed to code design and
manuscript pre-submission revision. Ranee Joshi implemented parts of the computer code and supporting algorithms for the
map2loop and contributed to manuscript pre-submission revision. Agnieszka Piechocka tested the code on different datasets
and contributed to code design and contributed to manuscript pre-submission revision. Lachlan Grose implemented parts of
the computer code and supporting algorithms for the *map2loop* and contributed to manuscript pre-submission revision. Miguel
de la Varga implemented parts of the computer code and supporting algorithms for the *map2loop* and contributed to manuscript
pre-submission revision. Laurent Ailleres contributed to code design and manuscript pre-submission revision. Guillaume Pirot
tested the code on different datasets and contributed to code design and manuscript pre-submission revision.

8. Acknowledgements

We acknowledge the support from the ARC-funded Loop: Enabling Stochastic 3D Geological Modelling consortia
(LP170100985) and DECRA (DE190100431). The work has been supported by the Mineral Exploration Cooperative Research
Centre whose activities are funded by the Australian Government's Cooperative Research Centre Programme. This is MinEx
CRC Document 2020/***. Source data provided by GSWA and Geoscience Australia.

600 **9. Appendices**

Appendix1. Minimum required inputs for *map2loop*

Minimum map2loop inputs:

605   1. EPSG coordinate reference system for input data (e.g. metre-based projection like UTM)

    2. Max/min coordinates of area of interest

    3. Geology polygons:

610

    -a. All polygons are watertight (node location mismatches must be within a smaller definable error)

    -b. Polygons have as attributes:

        -i. Object ID

        -ii. Stratigraphic code

615   -iii. Stratigraphic group

        -iv. One of more fields that describe if sill, if igneous, if volcanic

        -v. Min_age field

        -vi. Max_age field (can be same as Min_age field, and can be simple numerical ordering (bigger number is older))

620

    4. Fault/Fold Axial Trace Polylines:

    -a. Faults terminate on other faults but do not cross

    -b. Faults/Folds have as attributes:

        -i. Object ID

625   -ii. Field that determines if polyline is fault or fold axial trace

        -iii. Field that determine type of fold axial trace (e.g. syncline or anticline)

        -iv. Faults can have dip/dip direction info

    5. Bedding orientations:

630

    -a. Assumes dip/dip direction or dip/strike data

    -b. Orientations have as attributes:

        -i. Dip

        -ii. Dip Direction or strike

635

Appendix 2. Pseudocode for key calculations

640   **save_basal_contacts**

     explode geology polgyons so interior holes become distinct polygons
     for each polygon:
        build list of polygons and their atributes
645   load sorted stratigraphy from csv file
     for each polygon in list:
      if not intrusive:
        if polygon Code found in sorted stratigraphy:
          for each polygon in list:
650   if two polygons are not the same:
            if two polygons are neighbours:
              if second polygon is not a sill:
                add neigbour to list
        if first polygon has neighbours:
655   for each neighbour:
            if neighbour polygon Code found in sorted stratigraphy:
              if neighbour older than first polygon:
                calculate intersection of two polygons:
                  if intersection is a multilinestring:
660   for all line segments in linestring:
                      save out segment with x,y,z Code
                      build dictionary of basal contacts and dictionary of decimated basal contacts

     return dictionary of basal contacts and dictionary of decimated basal contacts
665

     save_basal_no_faults

     load fault linestrings as GeoDataBase
     create polygonal buffer aorund all faults
670   clip basal contacts to polygonal buffer
     make copy of clipped contacts
     for each clipped basal contact polyline:
      if polyline is GEOMETRYCOLLECTION:
        remove from copy of clipped basal contacts
675   else:
        add to dictionary

     build GeoDataFrame from remaining clipped basal contacts and save out as shapefile

680   **save_fold_axial_traces_orientations**

     load geology polygons as GeoDataFrame
     load interpolated contacts as array
     load polylines as GeoDataFrame
685   for each polyline:
      for each line segment in polyline:
        if fold axial trace:
          if passes decimate test:
            calculate azimuth of line segment
690   calculate points either side of line segment

find closest interpolated contact

 if interpolated contact is sub-parallel to fold axial trace:

  save orientation data either side of segment and related x,y,z,Code to csv file

695 **interpolate_contacts**

 create grid of positions for interpolation, or use predefined list of points

 for each linestring from basal contacts:

  if passes decimation test:

700    for each line segment in linestring:

    calculate direction cosines of line segment and save to file as csv with x,y,z,etc

 interpolate direction cosines of contact segments

705  save interpolated contacts to csv files as direction cosines and azimuth info with x,y,z,etc

interpolate_orientations

 subset points to those wanted

710  create grid of positions for interpolation, or use predefined list of points

 for each point from orientations:

  calculate direction cosines of orientations

 interpolate direction cosines of orientations

715

 save interpolated orientations to csv files as direction cosines and dip,azimuth info with x,y,z,etc

join_contacts_and_orientations

720  for each orientation in grid:

  rescale contact direction cosines with z cosine of orientations

  save out rescaled x,y direction cosines from contacts with z direction cosine from orientations and positional x,y,z,Code

calc_thickness

725

 load basal contacts as vectors from csv file

 load interpolated bedding orientations from csv file

 load basal contacts as geopandas GeoDataFrame of polylines

 load sorted stratigraphy from csv file

730  calculate distance matrix of all orientations to all contacts

 for each contact line segment:

  if orientations within buffer range to contact:

   calculate average of all orientation direction cosines within range

735   calculate line normal to contact and intersecting its mid-point

  for all basal contact polylines:

   if polyline Group is one stratigraphically one unit higher:

    if contact normal line intersects polyline:

     if distance between intersection and contact mid-point less than 2 x buffer:

740       store info

  from list of possible intersections, select one closest to contact mid-point

  if closest is less than maximumum allowed thickness:

   save thickness and location to csv file

9. Code availability

MIT Licence

http://doi.org/10.5281/zenodo.4288476

10. Data availability

http://doi.org/10.5281/zenodo.4288476

11. References

Ailleres, L., Betts, P., 1998. Geometrical and geophysical modelling of an inverted Middle Proterozoic fault system, Mount Isa Terrain, Australia. Conference Abstracts: 3D Modelling of Natural Objects: a Challenge for the 2000, Vol. 2. Nancy, France, 4–5th June.

Ailleres, L., Grose, L., Laurent, G., Armit, R., Jessell, M., Caumon, G., de Kemp, E., Wellmann, F., 2018. Loop – a new open source platform for 3D geo-structural simulations. Three-dimensional geological mapping workshop, Resources for Future Generations meeting, Vancouver, Canada, 14-18.

Aitken, A. R. A., Occhipinti, S. A., Lindsay, M. D. & Trench, A., 2018 A role for data richness mapping in exploration decision making. Ore Geology Reviews. 99,398-410.

Argand, 1911. Les nappes de recouvrement des Alpes Pennines et leur prolongement structuraux, Mat. carte géol. Suisse, N.S., XXXI livr.

Bond, C. E.: Uncertainty in structural interpretation: Lessons to be learnt, J. Struct. Geol., 74, 185–200, 2015

Bonham-Carter, G.F. 1994. Geographic Information Systems for Geoscientists, Pergamon, 398pp.

Brodaric, B., Fox, P., and McGuinness, D. L.: Geoscience knowledge representation in cyberinfrastructure, Computers & Geosciences, 35, 697-699, https://doi.org/10.1016/j.cageo.2009.01.001, 2009.

Burns, K.L., 1988, Lithologic Topology and Structural Vector Fields Applied to Subsurface Prediction in Geology, GIS/LIS'88, Proceedings 3rd Annual International Conference, Exhibits and Workshops, Volume 1, San Antonio, Texas, November 30-December 2, 1988, 26-34.

Cadman, W.K. 1927. The golden lanes of Greenwood County, Kansas, American Association of Petroleum Geologists , 11 (11): 1151-1172

Calcagno P, Chilès JP, Courrioux G, Guillen A (2008) Geological modelling from field data and geological knowledge: part I. Modelling method coupling 3D potential- field interpolation and geological rules, Phys Earth Planet Inter Vol. 171 (1) 147–157.

Carmichael, T. L., & Ailleres, L. (2016). Method and analysis for the upscaling of structural data. Journal of Structural Geology, 83, 121-133. https://doi.org/10.1016/j.jsg.2015.09.002

Caumon, G., Gray, G., Antoine, C., and Titeux, M.-O. 2013. Three-Dimensional Implicit Stratigraphic Model Building from Remote Sensing Data on Tetrahedral Meshes: Theory and Application to a Regional Model of La Popa Basin, NE Mexico. IEEE Transactions on geoscience and remote sensing, 51, 1613-1621

Cowan, E.J., Beatson, R.K., Ross, H.J., Fright, W.R., McLennan, T.J., Evans, T.R., Carr, J.C., Lane, R.G., Bright, D.V., Gillman, A.J., Oshust, P.A., and Titley, M., Practical implicit geological modeling, in Proc. 5th Int. Mining Conf.—Australian Inst. Mining Metallurgy, S. Dominy, Ed., 2003, pp. 89–99.

Giraud, J., Lindsay, M., Jessell, M. & Ogarko, V. 2020. Towards plausible lithological classification from geophysical inversion: Honouring geological principles in subsurface imaging. Solid Earth. 11, 419-436.

785   de Kemp, E.A., 1998. Three-dimensional projection of curvi-linear geological features through direction cosine interpolation of structural field observations. Computers & Geosciences 24 (3), 269–284.

de Kemp, E.A., Jessell, M.W., Aillères, L., Schetselaar, E.M., Hillier, M., Lindsay, M.D and Brodaric, B. 2017. Earth model construction in challenging geologic terrain: Designing workflows and algorithms that makes sense. Exploration 2017 Conference Paper.

790   de la Varga, M., and Wellmann, J.F., 2016, Structural geologic modelling as an inference problem: A Bayesian perspective: Interpretation, v. 4, p. SM1–SM16.

de la Varga, M., Schaaf, A., and Wellmann, F.: 2019. GemPy 1.0: open-source stochastic geological modelling and inversion, Geosci. Model Dev., 12, 1–32, https://doi.org/10.5194/gmd-12-1-2019

Farr, T. G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal,

795   D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D. 2007, The Shuttle Radar Topography Mission, Rev. Geophys., 45, RG2004, doi:10.1029/2005RG000183.

Fernández, O., 2005. Obtaining a best fitting plane through 3D georeferenced data. J. Struct. Geol. 27, 855–858. https://doi.org/10.1016/j.jsg.2004.12.004.

Freeman, L.C. 1977. A set of measures of centrality based on betweenness. Sociometry 40, 35-41.

800   Geological Survey of Western Australia 2016, 1:500 000 State interpreted bedrock geology of Western Australia, 2016: Geological Survey of Western Australia, digital data layer, <www.dmp.wa.gov.au/geoview>.

Geoscience Australia 2016, Digital Elevation Model (DEM) Shuttle Radar Topography Mission (SRTM) 1 Second over Australian Bathymetry Topography: Geoscience Australia, digital dataset, http://gaservices.ga.gov.au.

Geoscience Australia and Australian Stratigraphy Commission. 2017. Australian Stratigraphic Units Database.

805   https://asud.ga.gov.au/

Grose., L., Ailleres, L., Laurent, G., Armit, R., Jessell, M. 2019. Inversion of geological knowledge for fold geometry Journal of Structural Geology 119 (2019) 1–14

Hamilton, D.E., Didur, R.S., 1992. Three-dimensional geologic block modelling of the Kutcho Creek massive sulphide deposit, British Columbia. In: Hamilton, D.E., Jones, T.A. (Eds.), Computer Modelling of Geologic Surfaces and Volumes. AAPG

810   Computer Applications in Geology, No. 1, pp. 203–218.

Harrap, R. 2001. A Legend Language for Geologic Maps. Precambrian Times, 1, 1-9.

Houlding, S. W. (1994). 3D geoscience modelling computer techniques for geological characterization. Springer-Verlag. John Wiley & Sons Inc, New York, London, Sydney, Toronto.

Kelka, U, Westwerlund, S, Peeters, L. 2020. GIS based fault and fracture network analysis. Abstract, Sub 20 Conference,

815   Perth, Australia, 12-131th February, 2020. https://wp.csiro.au/sub20/program/

Lato, M.J., Vöge, M., 2012. Automated mapping of rock discontinuities in 3D lidar and photogrammetry models. Int. J. Rock Mech. Min. Sci. 54, 150–158. https://doi.org/10.1016/j.ijrmms.2012.06.003.

Lindsay, M D., Aillères, L., Jessell, M., de Kemp, E.A., and Betts, P.G., 2012, Locating and quantifying geological uncertainty in three-dimensional mod- els: Analysis of the Gippsland basin, southeastern Australia: Tectonophys- ics, v. 546–547, p. 10–

820   27.

Ma, X., and Fox, P.: Recent progress on geologic time ontologies and considerations for future works, Earth Science Informatics, 6, 31-46, 10.1007/s12145-013-0110-x, 2013.

Mallet, J. L. (2004). Space-Time Mathematical Framework for Sedimentary Geology. Mathematical Geology, 36(1), 1-32.

Mayoraz, R., Mann, C.E., Parriaux, A., 1992. Three-dimensional modelling of complex geological structures: new

825 development tools for creating 3-D volumes. In: Hamilton, D.E., Jones, T.A. (Eds.), Computer Modelling of Geologic Surfaces and Volumes. AAPG Computer Applications in Geology, No. 1, pp. 261–272.

NASA/JPL 2009. https://asterweb.jpl.nasa.gov/gdem.asp

Pakyuz-Charrier E, Giraud J, Ogarko V, et al (2018b) Drillhole uncertainty propagation for three-dimensional geological modelling using Monte Carlo. Tectonophysics. doi: 10.1016/j.tecto.2018.09.005

830 Pakyuz-Charrier, E., Lindsay, M., Ogarko, V., Giraud, J., and Jessell, M. 2018. Monte Carlo simulation for uncertainty estimation on structural data in implicit 3-D geological modelling, a guide for disturbance distribution selection and parameterization, Solid Earth, 9, 385–402, https://doi.org/10.5194/se-9-385-2018.

Perrin M, Rainaud J-F (2013) Shared Earth Modelling: Knowledge Driven Solutions for Building and Managing Subsurface 3D Geological Models

835 Ragan (1968) Ragan, D.M., 1968. Structural Geology: Introduction to Geometrical Techniques, first ed.

Ramsay, J.G., 1967, Folding and Fracturing of Rocks, McGraw-Hill, New York.

Rauch, A., Sartori, M., Rossi, E., Baland, P., Castelltort, S. 2109. Trace Information Extraction (TIE): A new approach to extract structural information from traces in geological maps. Journal of Structural Geology 126 (2019) 286–300

Sopwith, T. 1834. A Treatise on Isometrical Drawing as Applicable to Geological and Mining Plans, Picturesque Delineations

840 of Ornamental Grounds, Perspective Views and Working Plans of Buildings and Machinery, and to General Purposes of Civil Engineering. John Weald, London.

Thapa, P.B., Hoppe, A., 2012. 3D modelling of geological features. Bulletin of Nepal Geological Society, 2012, Vol. 29, pp. 67-72.

Thiele, S. T., Jessell, M. W., Lindsay, M., Ogarko, V., Wellmann, F., & Pakyuz-Charrier, E. (2016). The topology of geology

845 1: Topological analysis. Jorunal of Structural Geology, 91, 27–38.

Thiele, S.T., Micklethwaite, S., Bourke, P., Verrall, M., Kovesi, P., 2015. Insights into the mechanics of en-échelon sigmoidal vein formation using ultra-high resolution photogrammetry and computed tomography. J. Struct. Geol. 77, 27–44. https://doi.org/10.1016/j.jsg.2015.05.006.

Varnes, D.J. 1974. The Logic of Geological Maps, With Reference to Their Interpretation and Use for Engineering Purposes.

850 U.S. Geological Survey Professional Paper 837, 54pp.

Wellmann, F., Schaaf, A., de la Varga, M., & von Hagke, C. (2019). From Google Earth to 3D Geology Problem 2: Seeing Below the Surface of the Digital Earth. In Developments in Structural Geology and Tectonics (Vol. 5, pp. 189-204). Elsevier.

Wellmann, F., and Caumon, G. 2018. 3-D Structural geological models: Concepts, methods, and uncertainties Advances in Geophysics, Volume 59

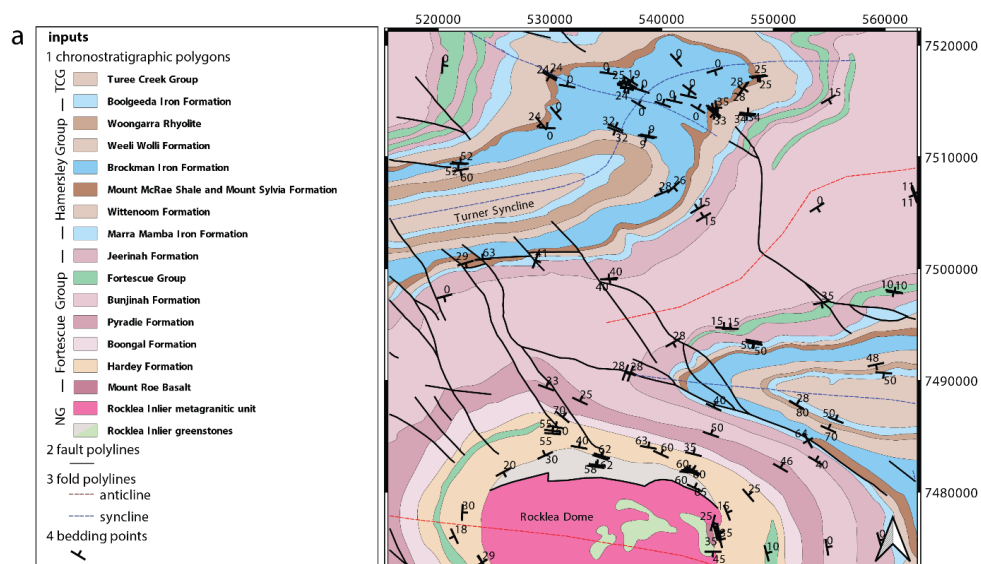
855 Wellmann, F., de la Varga, M., Murdie, R. E., Gessner, K., & Jessell, M. W. (2017). Uncer- tainty estimation for a geological model of the Sandstone greenstone belt, Western Australia—Insights from integrated geological and geophysical inversion in a Bayesian inference framework. Geological Society, London, Special Publications, 453, 41–52.

Wellmann, J.F., Lindsay, M., Poh, J., Jessell, M. 2014. Validating 3-D Structural Models with Geological Knowledge for meaningful Uncertainty Evaluations. Energy Procedia, 59, 374-381.

860 WFS standard, OGC

865

Table 1 Comparison between model engine inputs

Modelling Engine	Digital Terrain Model	Stratigraphy	Orientation data	Stratigraphic units	Faults	Fold axial traces
LoopStructural	Used	2-level	Bedding, Cleavages	Position, thickness of units	Position, age relationships wrt units and each other, displacement, ellipsoid for limited extent faults	Not used directly
GemPy	Used	2-level	Bedding	Position	Position, age relationships wrt units and each other, displacement, ellipsoid for limited extent faults	Not used directly

b **5 Stratigraphic (overlies ->) Relationships**

Turee Creek Group **->** Boolgeeda Iron Formation
Wittenoom Formation **->** Marra Mamba Iron Formation
Marra Mamba Iron Formation **->** Jeerinah Formation
Jeerinah Formation **->** Bunjinah Formation
Bunjinah Formation **->** Pyradie Formation
Pyradie Formation **->** Boongal Formation
Weeli Wolli Formation **->** Brockman Iron Formation

C 6 Digital Terrain Model

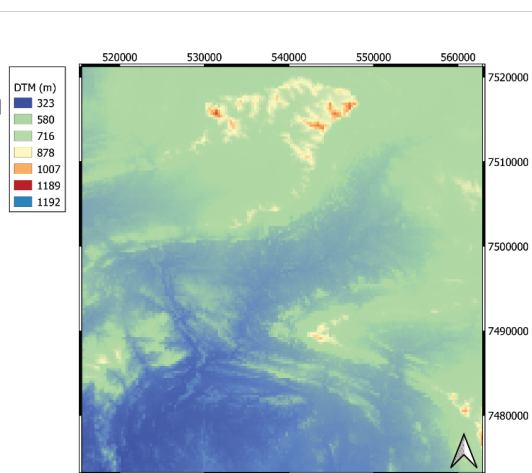



Figure 1. The six types of inputs to *map2loop*. **a) 1:500,000 Interpreted bedrock geology of the Rocklea Dome/Turner Syncline region of Western Australia showing the different datasets used to create the 3D model. TCG, Turner Creek Group. NG, no group defined by map, so each unit is its own group. The region shown is approximately defined by the max/min lat/long coordinates [117.15, -22.84, 117.60, -22.41]. b) Examples of the binary stratigraphic relationships derived from the Australian Stratigraphic Units Database (ASUD, Geoscience Australia and Australian Stratigraphy Commission. (2017). Australian Stratigraphic Units Database).** **c) The SRTM digital terrain model is sourced directly from Geoscience Australia at:** **http://services.ga.gov.au/gis/services/DEM_SRTM_1Second_over_Bathymetry_Topography/MapServer/WCSServer**

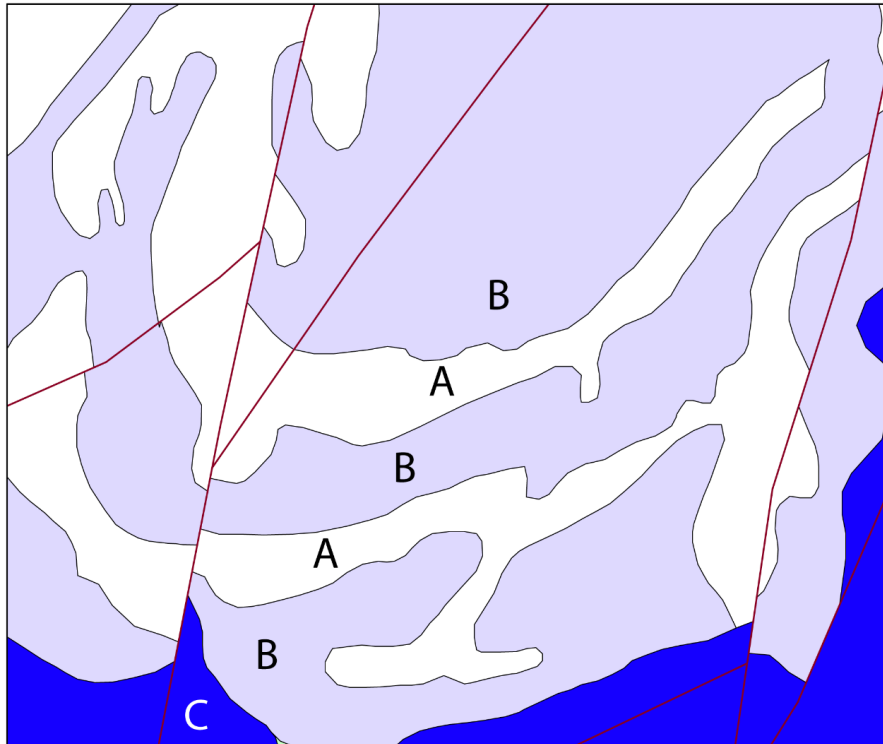
870

875

Figure 2. Example of lithological map descriptions that need recoding in order to work in a chronostratigraphic modelling workflow. Assuming that the repetition of units is not structurally controlled, the lithostratigraphic sequence C-B-A-B-A-B-A needs to be recoded as C-B1-A1-B2-A2-B3-A3.

880

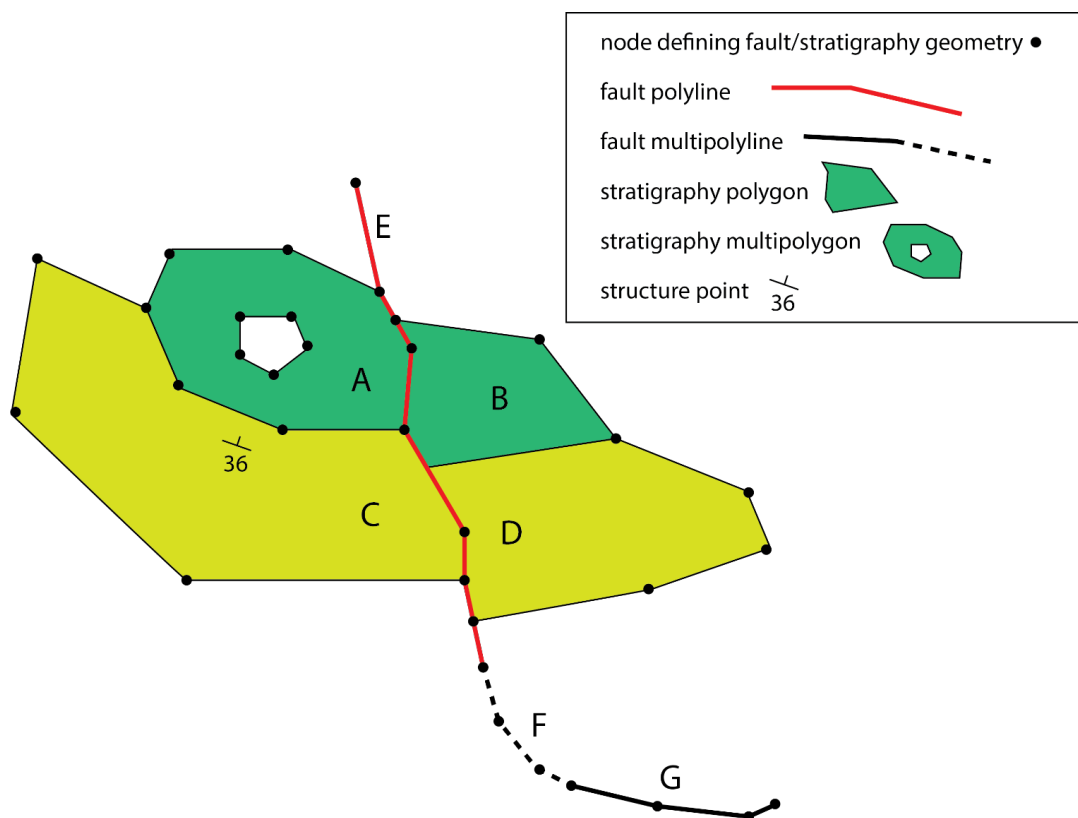



Figure 3. Geometric elements used in geological maps. B, C & D are stratigraphic POLYGONS, defined by a sequence of the x,y locations of nodes. A is a MULTIPOLYGON as it contains a hole, although MULTIPOLYGONS can also describe two unconnected
885 **polygons. E is a fault POLYLINE. F & G are fault MULTIPOLYLINES that describe segments of the same fault (as does fault E in this case). The structure observation (bedding measurement) is of type POINT. All geometric elements may possess multiple attributes, and are converted to 3D equivalents by add the information from the DTM.**

890

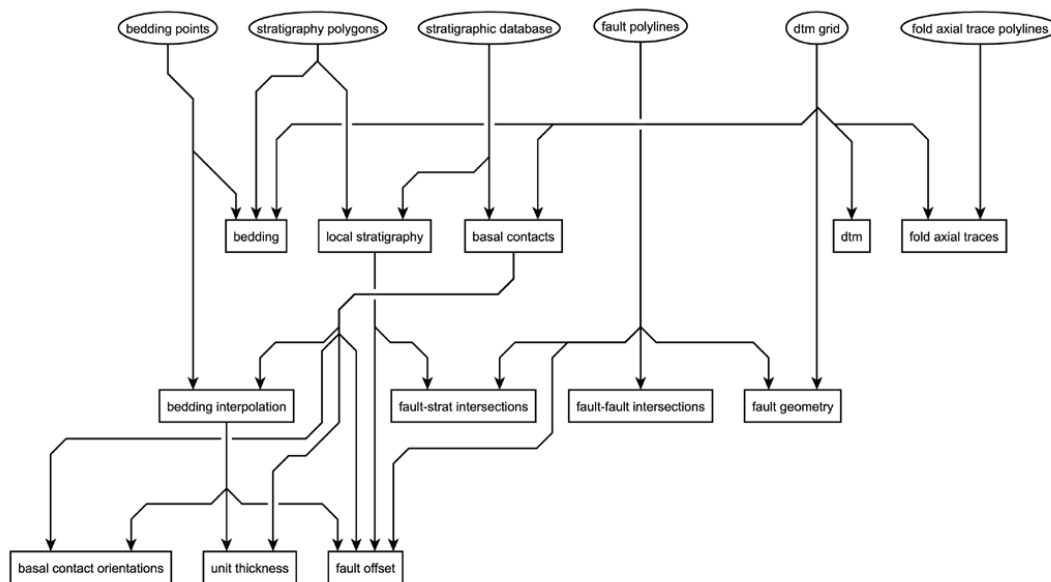



Figure 4. Data flow from inputs (ellipses) provided by GIS map layers, web servers, and stratigraphic databases. Augmented data (rectangles) are calculated by combining the inputs directly or incrementally during the *map2loop* workflow.

895

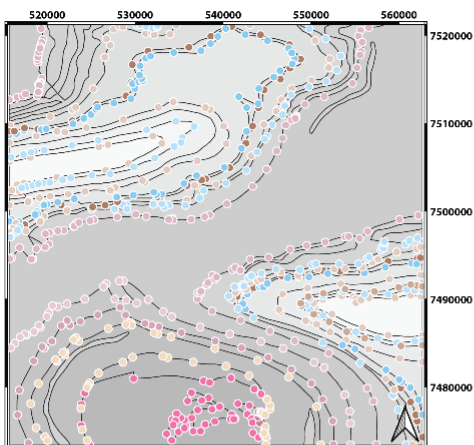
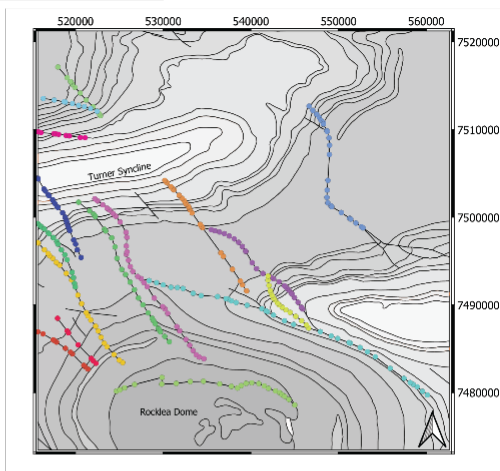
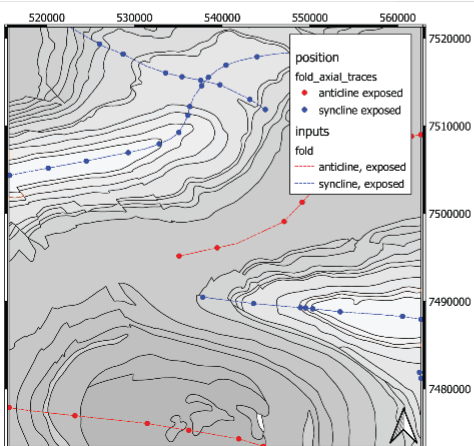
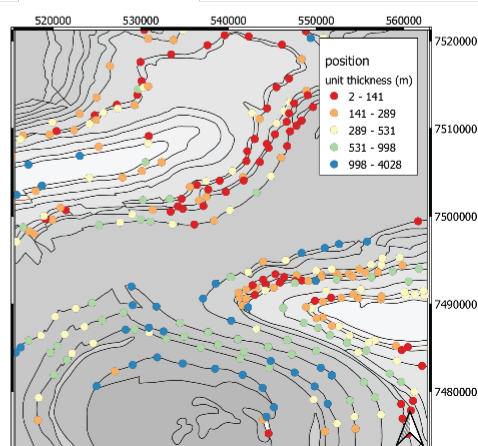
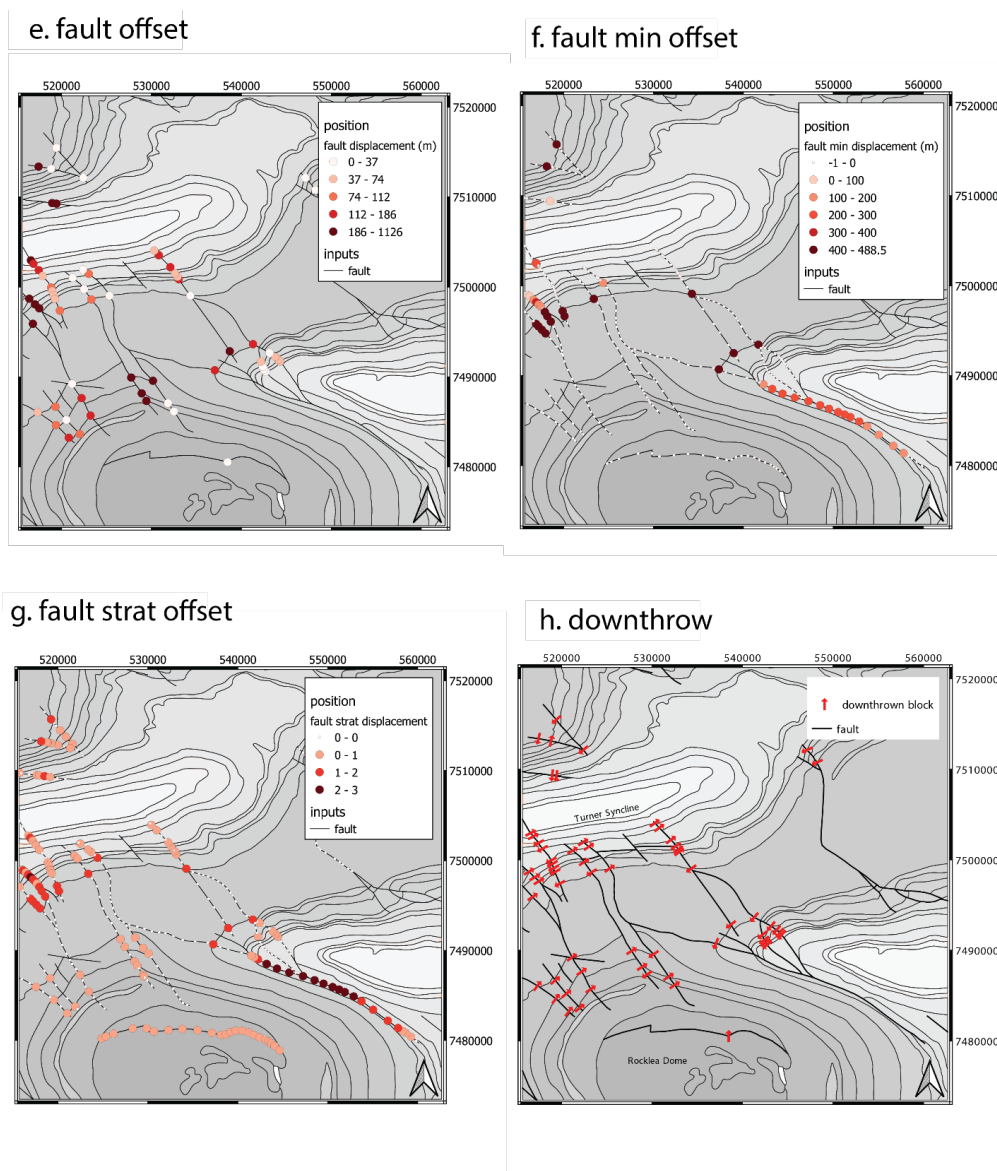



Figure 5. Positional information derived from map: a) Basal contacts of stratigraphic units, colours as Figure 1. b) Fault traces, colours randomly assigned to each fault, only faults longer than a defined length, in this case 5km, are processed. c) Fold axial traces, d) Local unit thicknesses. e) Fault offset, assuming down-dip displacement. f) Fault offset derived from minimum stratigraphic offset g) Stratigraphic fault offset and h) Fault downthrown block direction.

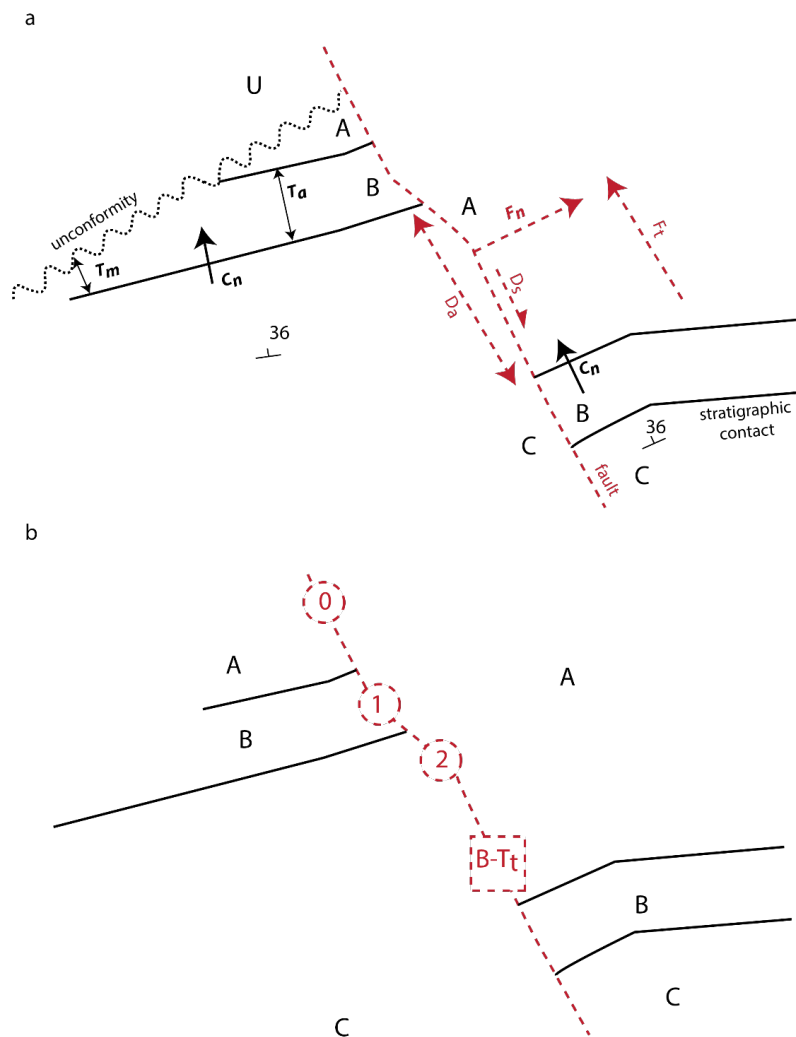



Figure 6. Positional calculations. a) Apparent unit thicknesses are calculated by calculating the normal distance from a contact (T_a), and are then transformed to 'true' thicknesses as by considering the local dip of the bedding. Apparent displacement is calculated by matching equivalent contacts across the fault, in this example the A-B contact (D_a). This is then transformed to 'true' displacement by assuming a down-dip slip vector. Finally the downthrown direction is calculated by examining the cross product of the fault trace (F_t) and the dip direction of the strata multiplied by the displacement. See text for details. b) If the direct calculation of fault displacement is not possible, because equivalent contacts across the fault cannot be established, then a minimum displacement can be estimated by the stratigraphic offset in terms of unit thicknesses. In the example here, the dashed red square indicates that the fault locally separates units A and C, so the minimum displacement is the thickness of unit B, which we were able to calculate above. If the unit thickness is not calculable for some reason, the stratigraphic offset between units A-A, A-B and A-C indicate a stratigraphic offset of 0, 1 and 2 stratigraphic units (red dashed circles).

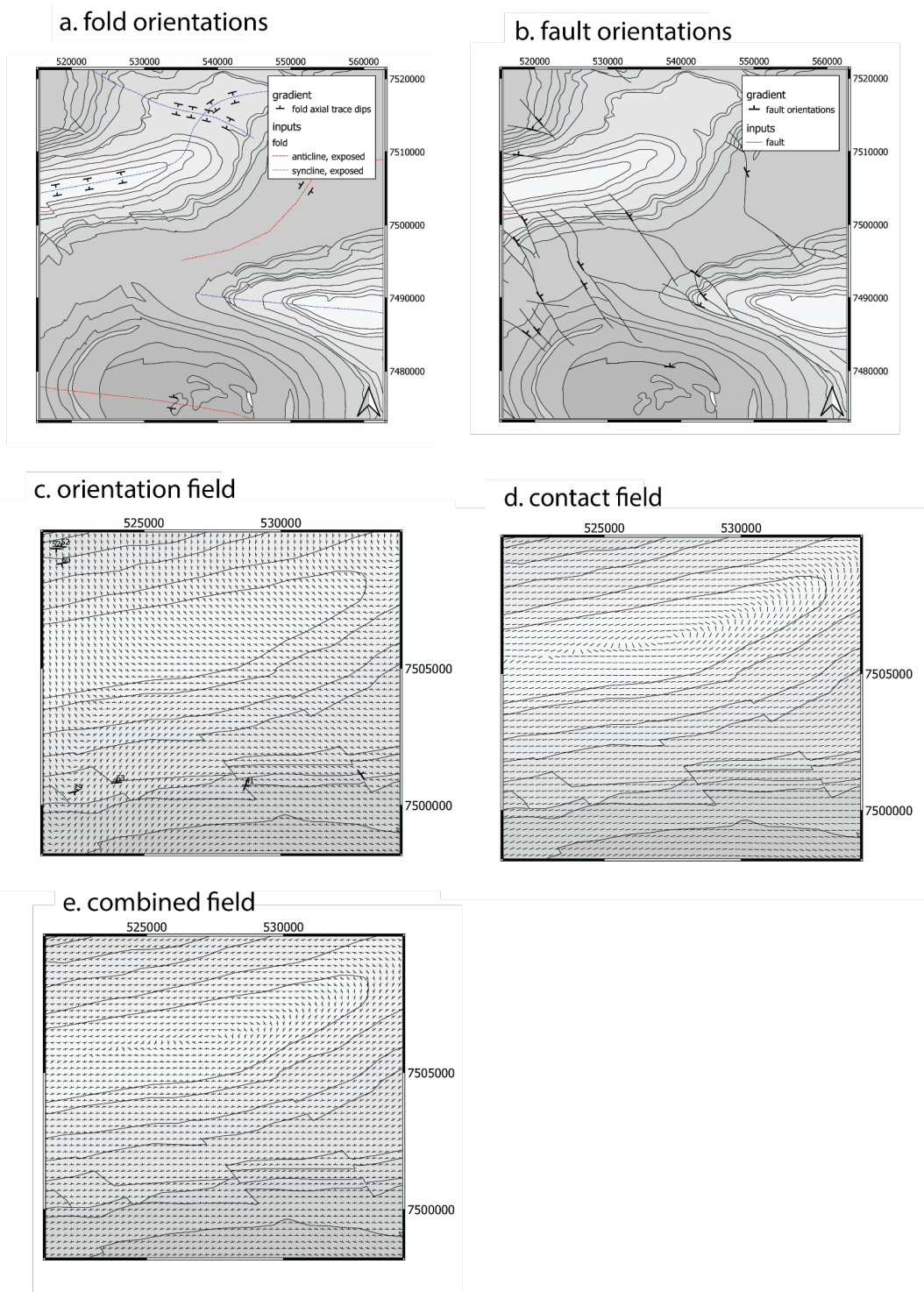



Figure 7. Gradient information derived from map: a) Bedding orientations near fold axial traces. b) Fault orientations. c) Interpolated orientation data, calculated as interpolated l_c,m_c , inset of part of NW area of map. d) Interpolated contact tangents, calculated as interpolated l_o,m_o,n_o direction cosines, inset of part of NW area. e) Combined information from interpolated dips and interpolated contacts, inset of part of NW area.

915

A->B : A stratigraphically overlies B

A-f-B : A is in faulted contact with B

———————▶ stratigraphically overlies

– – – – ▶ stratigraphically overlies and is also in faulted contact with

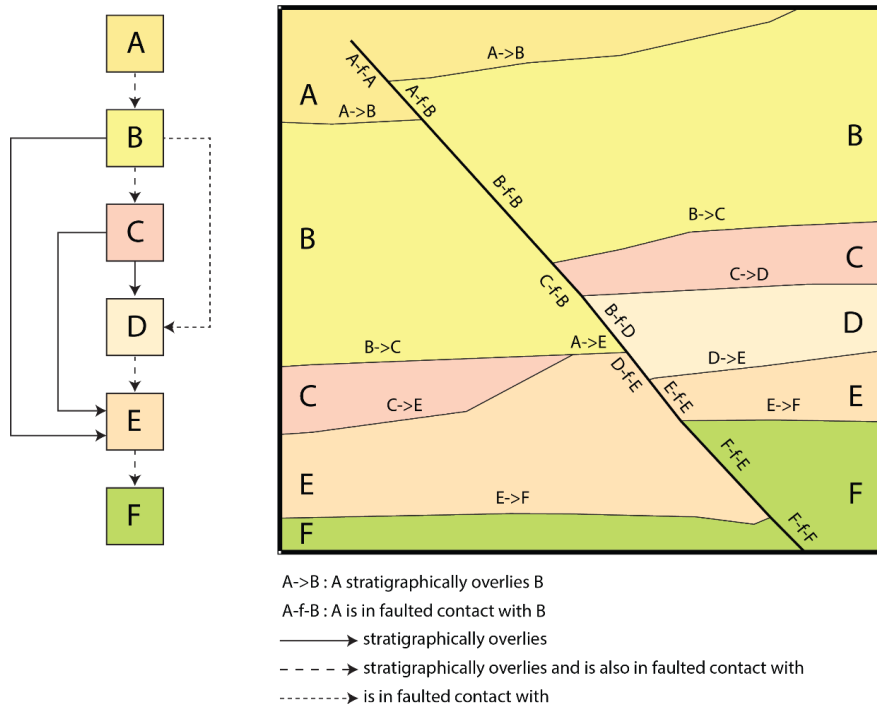
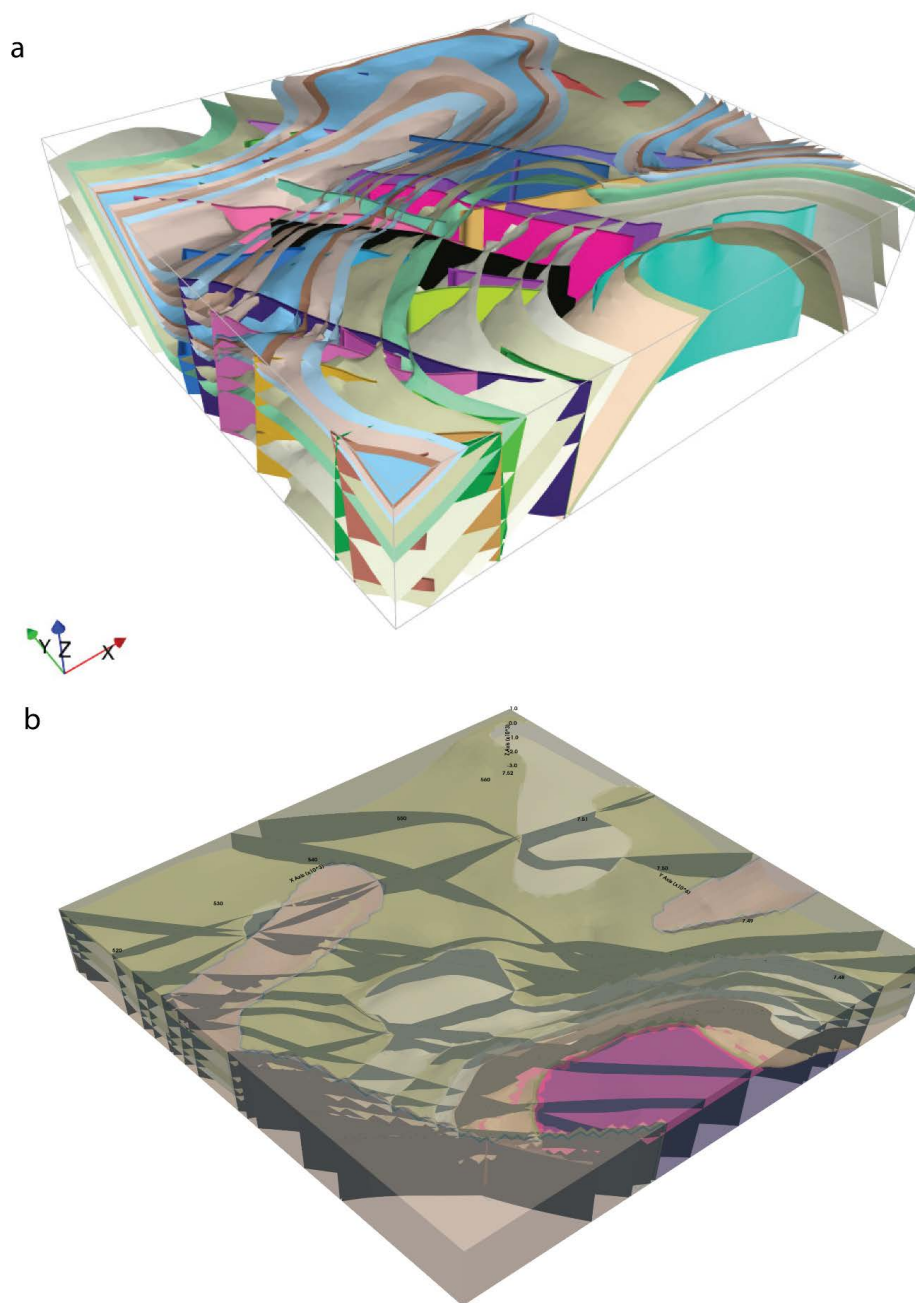
·········▶ is in faulted contact with

920

Figure 8. Example topological relations extracted from the map by the *map2model* library. In this map we have 6 units A-F which locally are in contact with each other either by normal stratigraphic relationships (A->B signifying that A is younger), or separated by a fault (A-f-B with no relative age significance). Once these individual binary relationships are aggregated by *map2model* into a single graph, specific pairs of units may be stratigraphic only (solid line), a combination of stratigraphic and fault relationships (long dashed line) or fault-only (short dashed line). Intrusive relationships (not shown here) will also be extracted from the map where present. These relationships form the basis of our understanding of the local stratigraphic graph.

925

Figure 9. Topological information derived from map: a) Stratigraphic ages relationships extracted from map and ASUD. Arrows point to older unit. Thickness of arrows is proportional to contact length. b) Fault labels for faults longer than 5km, and resulting fault-intersection relationships. Arrows point to older fault. d) Subset of fault-unit truncation relationships, yellow cell means that unit is not cut by a given fault, green cell means that it is.

a

b

935 **Figure 10. 3D Models built by a) LoopStructural and b) GemPy using the augmented data provided by *map2loop*. Note that different packages use different subsets of the available data. GemPy calculates limited-extent faults but currently displays them as extending across the model area.**