Computers & Geosciences 56 (2013) 76–82

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Relationships between palaeogeography and opal occurrence in Australia: A data-mining approach

T.C.W. Landgrebe*, A. Merdith, A. Dutkiewicz, R.D. Müller

The University of Sydney, School of Geosciences, Madsen Building, NSW 2006 Sydney, Australia

A R T I C L E I N F O

Article history:
Received 6 November 2012
Received in revised form
6 February 2013
Accepted 9 February 2013
Available online 16 February 2013

Keywords:
Data-mining
Spatio-temporal analysis
Palaeogeography
Opal
Association rules
Great Artesian Basin

A B S T R A C T

Age-coded multi-layered geological datasets are becoming increasingly prevalent with the surge in open-access geodata, yet there are few methodologies for extracting geological information and knowledge from these data. We present a novel methodology, based on the open-source GPlates software in which age-coded digital palaeogeographic maps are used to "data-mine" spatio-temporal patterns related to the occurrence of Australian opal. Our aim is to test the concept that only a particular sequence of depositional/erosional environments may lead to conditions suitable for the formation of gem quality sedimentary opal. Time-varying geographic environment properties are extracted from a digital palaeogeographic dataset of the eastern Australian Great Artesian Basin (GAB) at 1036 opal localities. We obtain a total of 52 independent ordinal sequences sampling 19 time slices from the Early Cretaceous to the present-day. We find that 95% of the known opal deposits are tied to only 27 sequences all comprising fluvial and shallow marine depositional sequences followed by a prolonged phase of erosion. We then map the total area of the GAB that matches these 27 opal-specific sequences, resulting in an opal-prospective region of only about 10% of the total area of the basin. The key patterns underlying this association involve only a small number of key environmental transitions. We demonstrate that these key associations are generally absent at arbitrary locations in the basin. This new methodology allows for the simplification of a complex time-varying geological dataset into a single map view, enabling straightforward application for opal exploration and for future co-assessment with other datasets/geological criteria. This approach may help unravel the poorly understood opal formation process using an empirical spatio-temporal data-mining methodology and readily available datasets to aid hypothesis testing.

1. Introduction

Exploration practices and data analysis and modelling related to mineral and hydrocarbon systems are making increasing use of integrated geological datasets for understanding resource formation processes and thus improving exploration decision making. This is made possible by the rapid increase in the interconnected digital storage of data and the rapidly increasing power of computing systems. An important factor for many studies is the consideration of the palaeogeographic/depositional environment through time as inferred from the geological record, since the configuration of a particular region as observed today is nearly always substantially different from the geological past. This, together with an understanding of the physical processes involved, is particularly important for identifying niche geological conditions that result in a phenomena such as ore deposits, since only a consideration of the relevant geological history leads to an appropriate contextualisation of present-day observations. Thus investigating the nature and relationships involved in time-varying spatial data is a promising area for developing and applying machine learning to geological data. In this paper we develop a quantitative approach for utilising the palaeo-environmental history over a large portion of Australia to investigate some key factors associated with opal formation, as a step towards establishing more quantitative exploration criteria. A data-mining approach is taken here to cope with the large datasets involved, and to handle some degree of noise present in the datasets utilised, especially considering that the digital palaeogeographic maps this approach is based on involves interpretations and interpolations of sparse data points. The study is made possible through use of the GPlates plate tectonic geographic information system (Qin et al., 2012) and Gplates data-mining functionality (Landgrebe and Mueller, 2008), in which time-varying data are explicitly modelled, allowing for direct extraction of spatio-temporal associations.

---

* Corresponding author. Tel.: +61 2 9351 3625; fax: +61 2 9351 2442.
E-mail addresses: landgrebetcw@gmail.com,
thomas.landgrebe@sydney.edu.au (T.C.W. Landgrebe),
amer7632@uni.sydney.edu.au (A. Merdith),
adriana.dutkiewicz@sydney.edu.au (A. Dutkiewicz),
dietmar.muller@sydney.edu.au (R.D. Müller).

0098-3004/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2013.02.002

Opal is a form of hydrated silica ($SiO_2 \cdot n\ H_2O$) found predominantly throughout the Great Artesian Basin in Australia. Although the Australian opal fields are responsible for over 90% of the world's opal supply, there has been a decline over the past 20 years in the number of high quality gemstones produced by the fields (Smallwood, 1997). This is accounted for, to some extent, by the absence of formal exploration models for Australian opal, which has led to an over-reliance on mining old opal fields discovered in the early 20th century (Barnes and Townsend, 1990). Making use of known locations of opal, we construct a method allowing us to search for other locations that possess similar palaeogeographic sequences throughout the entire Great Artesian Basin, and help establish important time-varying palaeogeographic/depositional/facies conditions favourable for opal formation.

2. Eastern Australian opals and palaeogeography

Opal in the Great Artesian Basin is found within fractures and primary and secondary pore spaces in the upper 30 m throughout the highly weathered Cretaceous sedimentary sequence of the Eromanga and Surat basins (Barnes and Townsend, 1990). The location of opal deposits is shown in Fig. 1, and includes primary opal mining regions from which a total of 1036 mining localities were identified for this study. The locations were taken from maps published by the state geological surveys (Carr, 1979; Carter, 2005; Geological Survey of Queensland, 2012; Robertston and Scott, 1989), and were digitised and geo-referenced using the GIS software. The shallow stratigraphy of both basins consists of alternating layers of sandstones, claystones and siltstones deposited during the Early Cretaceous (ca. 135–95 Ma) (Campbell and Haig, 1999; Frakes et al., 1987). The generalised sequence consists of lower facies consistent with deposition in a cyclical marine environment followed by a marine regression leading to a fluvial–deltaic environment (Exon and Senior, 1976). Deposition ceased towards the end of the Cenomanian, and until the Palaeogene the area experienced uplift and erosion resulting in up to 3 km of sedimentary rock being eroded (Raza et al., 2009). Further sediment deposition initiated in the Palaeogene (Alley, 1998), as well as substantial periods of silicification of the outcropping sedimentary rocks due to weathering (Alley, 1998; Thiry et al., 2006). However, the relative tectonic stability of the Australian plate (Veevers, 2000) and the weathering history since the late Cretaceous has allowed for the (partial) preservation of the Cretaceous sedimentary sequences, either allowing for the preservation of opal or for the necessary constituents for the formation of opal (e.g. source of silica). Since Australia produces over 90% of precious sedimentary opal globally, this suggests that the palaeogeographic evolution of Australia is important for understanding how opal may have formed and for determining where it may be found.

The Australian palaeogeographic atlas dataset (Langford et al., 1995; Totterdell, 2002) is a digital version of a compilation of maps based on palaeo-reconstructions of the environment, latitude, and sedimentology of the Australian continent. The dataset distinguishes between 16 differing palaeo-environments (defined in Table 1) and provides a geological reconstruction back to 542 Ma, comprising 70 time slices. As opal is found predominantly in Cretaceous sedimentary rocks, the analysis of the dataset was constrained to the time period from 145 Ma to the present-day. In total this comprises 19 time slices, depicted in the time-slice maps in Fig. 2 (note that the figure involves only 9 of the 16 defined environments that pertain to these time slices).

Fig. 3 illustrates the various palaeogeographic time series profiles for major Australian opal-mining locations. This plot clearly shows that there are broad similarities in the palaeogeographic sequences across the various localities (i.e., there is a subset of profiles to explain the majority of the environmental evolution through time), in

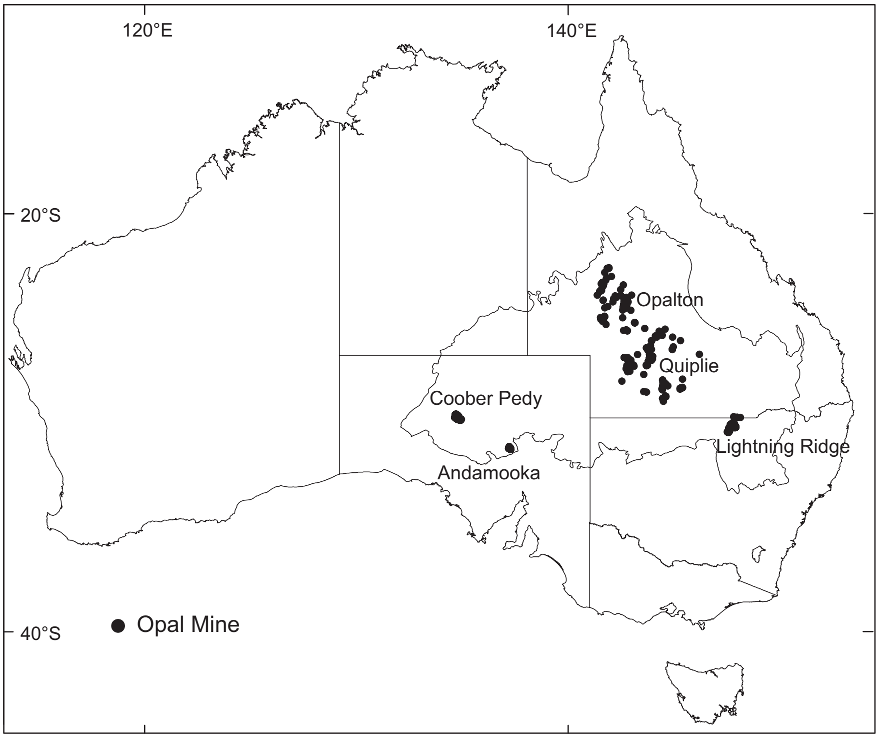

Fig. 1. Location map for this study, with the outlined Great Artesian Basin defining the broad prospective region in which most Australian opals are found. The locations of opal mining localities used in this study are shown.

particular a broad tendency for a transition between *land depositional fluvial* and *lacustrine* environments into a *shallow marine* environment, followed by land-based *fluvial*, *erosional* and *coastal paralic* environments. It is the translation of this raw representation into a compact, less “noisy” form applicable for generating a large-scale prospectivity map that is the principle step taken in this paper (see next section).

Table 1
Defined palaeogeographic environments used in this study, and their abbreviations. Colour-coding is as per Langford et al. (1995). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Palaeogeographic enviroment	*Abbreviation*
Coastal depositional deltaic	CoastDepDel
Coastal depositional intertidal-supratidal	CoastDepInterSup
Coastal depositional paralic	CoastDepPar
Land depositional aeolian	LandDepAeo
Land depositional fluvial	LandDepFluv
Land depositional fluvial-lacustrine	LandDepFluvLacu
Land depositional glacial	LandDepGlac
Land depositional lacustrine	LandDepLacu
Land depositional playa	LandDepPlaya
Land environment erosional	LandEnvEro
Land environment fluvial	LandEnvFluv
Marine abyssal	MarineAb
Marine bathyal-abyssal	MarineBathAb
Marine shallow	MarineSh
Marine shallow abyssal	MarineShAb
Marine very shallow	MarineVSh

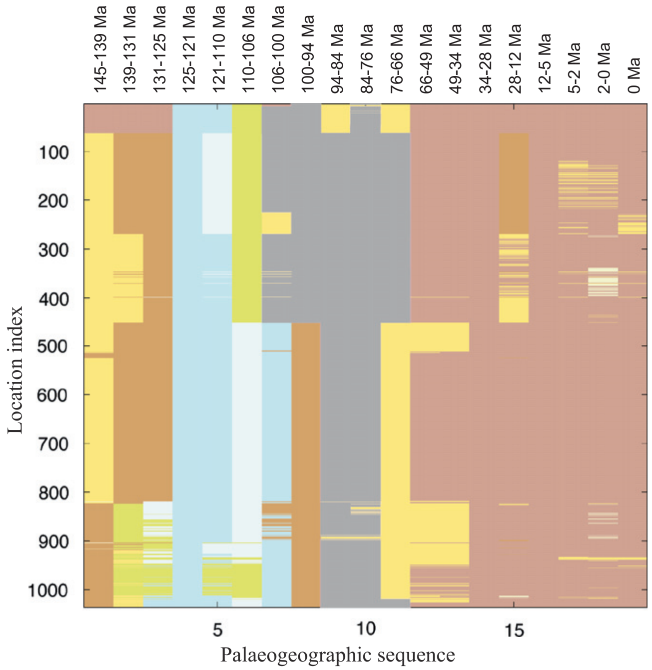

Fig. 3. Graphical depiction of the palaeogeographic time series at known opal mining locations (a total of 1036) computed across the 19 time-slices. The vertical axis indexes the various localities, and the horizontal axis depicts the time period per time slice in the atlas. Grey values indicate unknown or undetermined properties. The colour-scale is defined in Fig. 2.

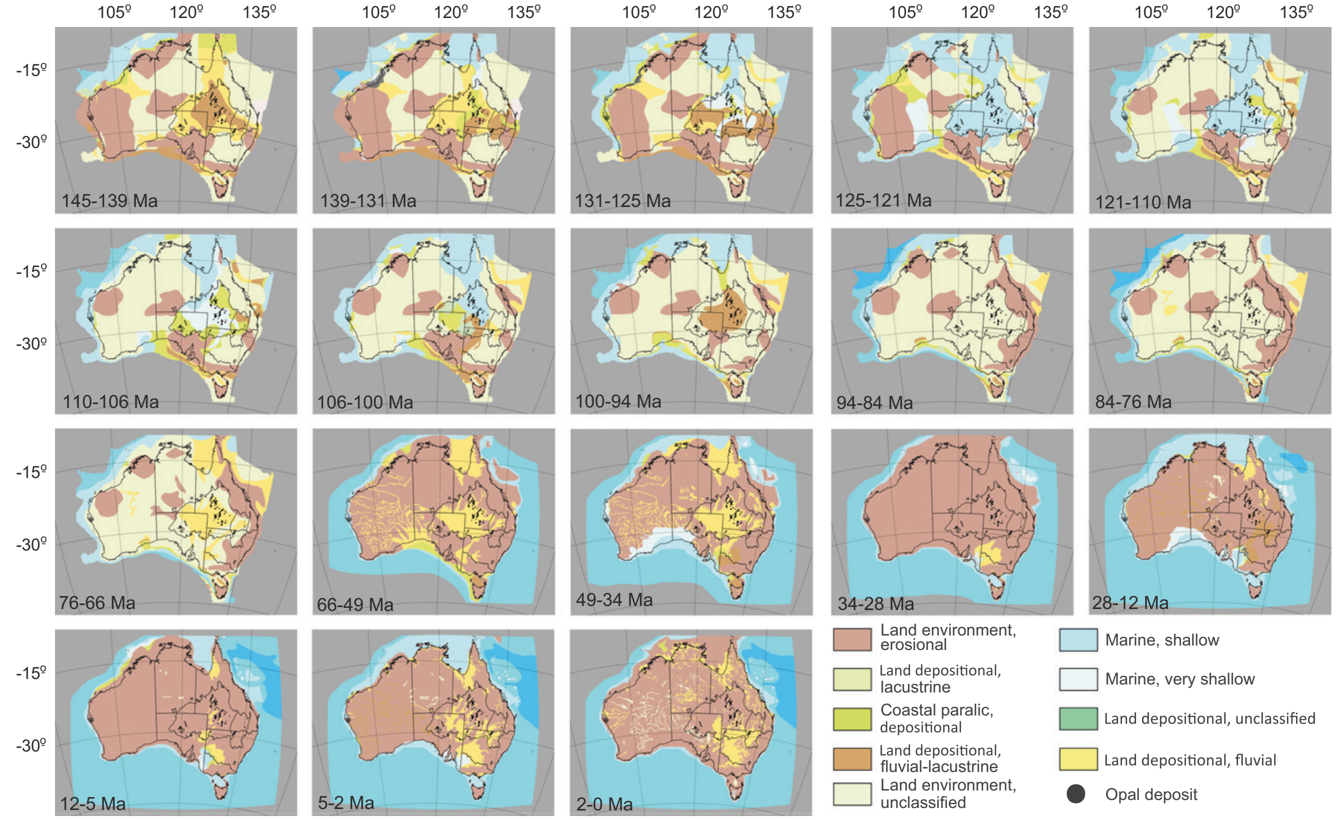

Fig. 2. Palaeogeographic atlas time slices for the period of interest, with the main opal localities indicated. These clearly show the dynamic nature of the geographic environment since the Cretaceous, with the general link between the occurrence of the Great Artesian Basin environment and the presence of opal. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Formalising the palaeogeographic time-series analysis

The quantitative methodology followed in this paper utilises the palaeogeographic dataset, referenced with known opal localities to derive a data-driven model of target palaeogeographic sequences. The digital representation consists of spatial zones defined by polygons, with associated geographical attributes. We define the unique set of 16 different palaeo-environment types $\vec{\omega}=[\omega_1,\omega_2,\ldots,\omega_{16}]$, where the *i*th type is referred to as ω_i (the environment types are listed in Table 1). In this study it is of interest to obtain time-series representations of depositional environments (defined as the "palaeogeography") at various test latitude–longitude localities (denoted (θ_k,λ_k) for a location *k*). Each (θ_k,λ_k) location is thus associated with a palaeogeographic sequence through the various time slices, with the geography at time t_i denoted as $\omega(\theta_k,\lambda_k,t_i)$. The geological ages of interest fall within the Cretaceous to present-day period, with the most known opal occurring within the Cretaceous sediments geological region in eastern Australia. This period constitutes a total of 19 time slices in the palaeogeographic atlas, denoted as $(t_1,t_2,\ldots,t_{19})$. A palaeogeographic time series (γ) at (θ_k,λ_k) is then defined as $\gamma(\theta_k,\lambda_k,\vec{t})=[\omega(\theta_k,\lambda_k,t_1),\omega(\theta_k,\lambda_k,t_2),\ldots,\omega(\theta_k,\lambda_k,t_{19})]$. The time series $\gamma(\theta_k,\lambda_k,\vec{t})$ is computed by identifying the polygon in which the test location resides at a given time slice. See Fig. 3 for a depiction of the various palaeogeographic γ time series profiles extracted.

In the next step, palaeogeographic time series are converted into ordinal event chronologies by removing duplicate neighbouring values through time, thus mapping the $\gamma(\theta_k,\lambda_k,\vec{t})$ time series into ordinal sequences denoted $\gamma_o(\theta_k,\lambda_k)$, where γ_o has a variable length ≤ 19 dependent on the number of event transitions. These representations are thus devoid of absolute time, representing only the chronological order of events. These event sequences are computed by first constructing a binary indicator function $I(\vec{t})$ (with length of $\vec{t}$) to determine transitions between events as follows:

$$I(t_i)=\begin{cases}1 & \text{if } i=1\\ 1 & \text{if } i>1 \quad \text{and} \quad \gamma(\theta_k,\lambda_k,t_i)\neq\gamma(\theta_k,\lambda_k,t_{i-1})\\ 0 & \text{otherwise}\end{cases} \tag{1}$$

The ordinal sequences can then be computed as

$$\gamma_o(\theta_k,\lambda_k,t_i)=\begin{cases}\gamma(\theta_k,\lambda_k,t_i) & \text{if } I(t_i)=1\\ empty & \text{otherwise}\end{cases}\forall i,\ i=[1,2,\ldots,19] \tag{2}$$

Some undefined environments are present in the atlas, and tend to cover large regions. In our analysis these are removed from the targeted event sequence list, recognising that this would affect the pattern matching methodology. The assumption made is that since the undefined sequences typically occupy broad regions, the removal will affect both opal locations and arbitrary locations in the same way, with the resultant pattern matching remaining robust.

4. Event sequence pattern-matching

Eq. (2) provides the means to represent the nature of the palaeogeographic evolution at a particular location without requiring absolute ages. This is important for opal exploration, since direct radiometric dating of Australian opal has not been possible due to its low uranium content (Gaillou et al., 2008). Consequently, it has not been possible to constrain the age of Australian opal further than the age of its host rock, which is predominantly Cretaceous, with a single location (Mintabie) associated with Ordovician sandstones. By removing the requirement for absolute time, our analysis accommodates an age range for opal from the Cretaceous to the present day. The primary control being exploited is thus a favourable evolution and juxtaposition of geological environments.

We extract a dataset consisting of palaeogeographic time series from the 1036 identified opal localities (Fig. 3), and convert these into the event sequences via Eq. (2). Amalgamating the event sequences yields an interesting result, namely that there is considerable similarity between the sequences across all opal localities, a result that may help validate the fundamental conjecture in this study—that the environmental event chronology is a primary control on opal formation. Of the 1036 opal localities, 52 independent ordinal sequences were identified, and of these, only 27 sequences were found to describe 95% of all independent ordinal palaeogeographic sequences. We make the assumption that these 27 sequences are representative of the palaeogeographic sequences responsible for opal formation, discounting the remaining 25 more spurious ordinal sequences. We must note that the sampling of the opal localities is somewhat biased by issues such as legislation (Australian State laws limit extensive opal mining at certain locations) and exploration bias, but argue that by retaining 95% of the target ordinal sequences, the resulting set should capture the broad nature of the favourable palaeogeography.

Subsequent to the sequence filtering step, the resulting 27 ordinal sequences are depicted in Fig. 4, providing a visual impression of the palaeogeographic event sequences. Note that the length of each target ordinal sequence varies arbitrarily according to the number of event transitions, which is handled next by the pattern matching algorithm. These broadly suggest that the initial depositional environment is land-based fluvial lacustrine and fluvial. A transition then generally occurs to land-based fluvial and coastal paralic environment, and then to shallow or very shallow marine environments. This is followed by transitions to coastal or land-based fluvial lacustrine, then fluvial, and finally a general transition to erosional environments.

The target event chronological information encapsulated in Fig. 4 is treated as a *training set* (defined as $\Gamma_o^T=[\gamma_{o_1}^T,\gamma_{o_2}^T,\ldots,\gamma_{o_M}^T]$ for *M* ordinal sequences) for an automated palaeogeographic

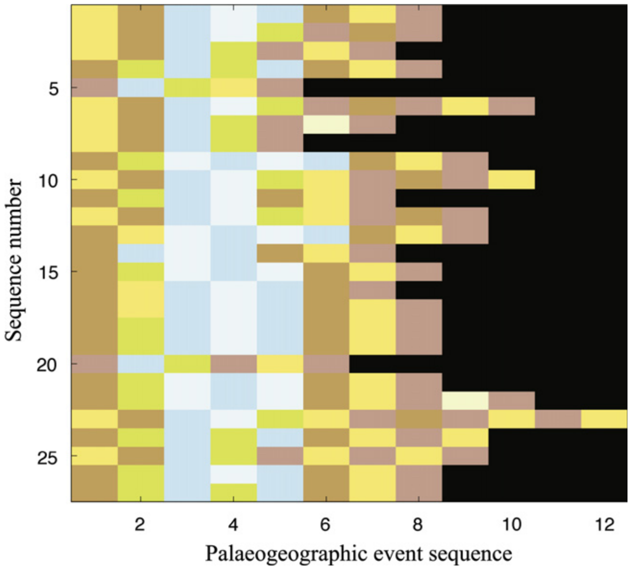

Fig. 4. Extracted ordinal sequences for time-series originating from the training phase. This is a similar representation to Fig. 3, except that horizontal axis depicts the event transition chronologies only (i.e., absolute time is factored out). The colour-scale is defined in Fig. 2, with the black regions representing the end of the sequences. These 27 sequences represent the primary transition sequences that describe 95% of opal occurrences.

prospectivity analysis for eastern Australia. The approach taken is to consider the palaeogeographic event chronologies at each spatial location (denoted (θ_k,λ_k) for an arbitrary location) in the target region, and match these with the training sequences. A grid of test locations is generated across eastern Australia with a 200 m spacing (chosen to match available high-resolution geophysical datasets to be analysed in conjunction with this study, as published in Merdith et al. (in press). The bounding latitude–longitude extent is as follows: [(17.8604°S, 131.5106°E), (32.6709°S, 151.8217°E)], covering the Great Artesian Basin and surrounds, with the resulting point data used to define a map layer suitable for large-scale prospectivity mapping. Each position on this map is coded with either a 0 or 1 according to its match to the palaeogeographic training set. If the ordinal event sequence at an arbitrary location (θ_k,λ_k) is defined as $\gamma_o(\theta_k,\lambda_k)$, the output classification $f(\gamma_o(\theta_k,\lambda_k,\mathbf{t}))$ (where $\mathbf{t}$ represents the length of that ordinal sequence) is defined as follows:

$$f(\gamma_o(\theta_k,\lambda_k,\mathbf{t})) = \begin{cases} 1 & \text{if } \gamma_o(\theta_k,\lambda_k,\mathbf{t}) = \gamma^T_{o_i} \ \forall i, \quad i=1,2,\ldots,M \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

In this way, each location in the area of the case study can be matched to the trained event sequences, and spatial regions formed indicating locations favourable for prospectivity.

5. Computing methodology

The plate-tectonic Palaeogeographic Information System *GPlates* was used as the basis for this study (Boyden et al., 2011). GPlates provides the ability to assess multiple spatial datasets that include time-varying properties, allowing for visualisations and analyses coherent both in space and time. Thus factors such as a varying geographic environment, or spatial relationships between different datasets undergoing plate motions or crustal deformation, can be assessed. GPlates has been augmented with a quantitative data-association tool (called the “coregistration tool”), and interfaced with the data-mining software infrastructure *AILab Orange* (Demsar et al., 2004). These technologies allow predefined relationships between datasets to be specified, typically resulting in time-series, that can then be quantitatively analysed for more detailed investigations.[1]

The first step in the computational methodology involves the spatio-temporal analysis between two spatial datasets, namely the opal locations and the palaeogeography. These datasets were loaded into GPlates, and the analysis was configured to run between 145 Ma and present day, comprising the period of interest for the study. The coregistration tool was used to compute the association between opal locations, and the “ENVIRONMENT” property in the palaeogeography dataset, thus resulting in a time series for each location representing the palaeogeographic environment at different time instances. The data mining environment was then used to assess the various time series, making use of a component called “OrdinalSignature” to convert the time series into event sequences. Similarly, 10,000 randomly chosen locations (uniformly generated on a sphere) were used to extract ordinal sequences at arbitrary locations. This resulted in a convenient data representation, allowing the various associations to be computed, and the target sequences to be extracted (Sections 4 and 6.1).

The second step in the computational methodology involves investigating the targeted associations across the broader eastern Australian region to identify regions with the same “signatures”. A 200 m spacing was used to generate a grid of latitude–longitude locations in the region of interest, followed by extraction of the time-varying signatures and ordinal representations (some use of the software tool *Matlab* was made to speed up this large calculation). These were matched to target signatures to create a match score as depicted in Fig. 5.

6. Results and validation

The result of the pattern matching methodology is shown in Fig. 5, with the Cretaceous sedimentary rocks highlighted, and filled black regions depicting areas matching the trained palaeogeographic patterns (targeted regions). Two general observations can be made. The first observation is that there is a strong correlation between the region encompassing Cretaceous sedimentary rock and the targeted regions. This is an expected result since this region has a unique palaeographic characteristic compared to neighbouring regions, and all the opal localities used for training were extracted from it. A second observation is that there is a substantial reduction in area, which is computed to consist of only approximately 11% of the original area, implying that the particular palaeogeographic sequences used in the training phase are somewhat distinct from those characterising the majority of the Great Artesian Basin. This output map layer could aid in future exploration, especially when considered alongside other geological and geophysical criteria (see also Merdith et al., in press).

6.1. *Data mining association rules*

The palaeogeographic sequence matching results show that only about a 11% of the Great Artesian Basin experienced the same palaeogeographic sequences as the 1036 known opal locations. A question is whether a specific set of key sequences is essential for opal formation. This information can be extracted from the computed ordinal signatures by looking for “key” event transitions, which is a standard practice in data mining, where frequently used “items” in a dataset collection are extracted on the basis of consistency/frequency, known as the data-mining of association rules (Tan et al., 2006). Since we are studying *ordinal* sequences, an additional constraint to be used is the temporal ordering, resulting in *sequential* association rules. These are computed by analysing both the target sequences, as well as sequences extracted from 10,000 randomly selected locations within the study area (with the unique set computed to consist of 1613 sequences, Fig. 6).

Two sets of association-rule data-mining results are established, as tabulated in Table 2. The relative itemset frequency (i.e., the abundance of that particular association rule) for the targeted sequences (which are extracted from known opal locations) are compared with those from the random set. In the first set, it can be seen that prior to a progression towards a marine environment, most of the targeted sequences experience a transition from a “LandDepFluv” to a “LandDepFluvLacu” environment (37%) or a transition from a “LandDepFluvLacu” to a “CoastDepPar” environment (40.7%). The strength of this association is ascertained by comparing these with random locations, where only 11.7% and 0.6% of samples respectively undergo the same transition. In the second set, we investigate the environment transition after the first “Marine” or “MarineSh” event, and calculate that all the target sequences undergo a transition to either “LandDepFluvLacu” or “CoastDepPar”, whereas in the random case, this association related only to 9.4% and 13.8%, respectively, of the samples. This suggests a second important association “rule” that is very specific to opal formation. The final

---

[1] Introductory material describing these capabilities can be found at https://sites.google.com/site/gplatestutorials/, see Data Mining tutorials (last accessed on 25 January 2013).

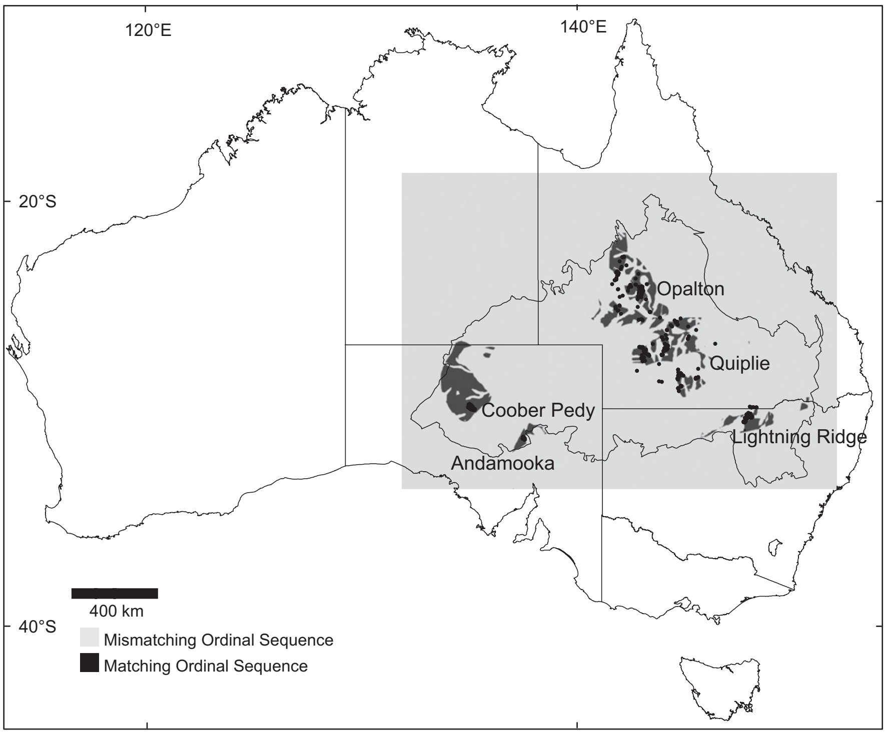

Fig. 5. Prospectivity map at 200 m resolution resulting from the time-series pattern matching. The light shaded region is the broad region over which the automated analysis was undertaken, encapsulating the inner polygonal region representing the Australian Cretaceous sedimentary sequence. The dark regions show all locations in which the pattern matching classified as target, also demonstrating that these are constrained to the Great Artesian Basin region as expected.

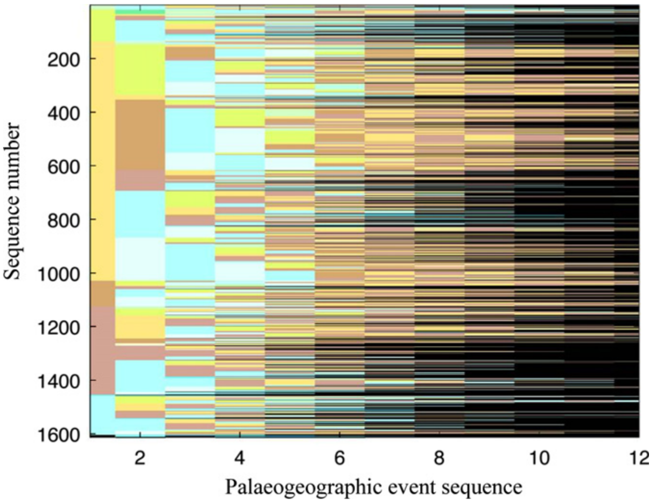

Fig. 6. Extracted ordinal palaeogeographic sequences as in Fig. 4 for the unique set of 1613 of 10,000 randomly selected locations. The colour-scale is defined in Fig. 2.

Table 2
Association rules data-mined from the 27 target sequences. Two "sets" of association rules are computed, namely (*Set1*) the events prior to the first Marine event, and (*Set2* the transition between Marine environments and land environments. Item-set frequency is the number of "items" (i.e., allowed sequences in the set following that particular rule).

Palaeogeographic sequences	*Item-set frequency (opal locations)*	*Item-set frequency (random locations)*
Set1		
LandDepFluv ⇒ LandDepFluvLacu	10/27 = 37.0%	1170/10,000 = 11.7%
LandDepFluvLacu ⇒ CoastDepPar	11/27 = 40.7%	61/10,000 = 0.6%
LandDepFluvLacu ⇒ LandDepFluv	3/27 = 11.1%	6/10,000 = 0.1%
Set2		
Marine ⇒ LandDepFluvLacu	13/27 = 48.1%	944/10,000 = 9.4%
Marine ⇒ CoastDepPar	14/27 = 51.9%	1384/10,000 = 13.8%

association rule is based on the observation that many of the random locations involve multiple transitions between marine and land environments. We find that 19.5% of the randomly selected samples involve these multiple events, which is generally not the case for the target locations.

A third association that appears prominent relates to the observation that 24/27 of the targeted sequences involve a single "marine" event. In other words, there is only a single palaeogeographic occurrence of transitions from land environments, marine environments, and back to land environments. The remaining three targeted sequences consistently allow for transitions between marine environments, coastal-depositional paralic environments, and back to marine environments, with a consistent transition to a land-depositional fluvial–lacustrine environment thereafter.

6.2. Discussion

The pattern matching methodology has provided the means to simplify a complex record of time-varying geology into a single map-view targeted specifically for opal occurrence. Such reductions in complexity are important for resource exploration where both the diversity and amount of geological datasets is growing rapidly, and practical methods are required to aid in effective, repeatable analysis. A similar procedure could be carried out for other ore deposits in which relevant spatio-temporal datasets are available, such as using ferruginous regolith material as a tracer for gold exploration (Anand, 2001). The association-rules analysis highlighted a number of geological environment transitions that are relatively specific for opals as compared to arbitrary locations. This study shows that a quantitative approach is able to systematically extract these types of scientific insights, as compared to a visual approach that is limited in coping with the large dataset sizes and time-varying nature.

7. Conclusions

This paper presents a quantitative methodology for utilising age-coded geological data to develop models/targeting criteria for opal exploration. Such data-driven approaches are becoming increasingly viable as the growth in the storage of digital geological data continues to accelerate, allowing for repeatable and less-subjective analyses. A palaeogeographic dataset consisting of age-coded geological environments in Australia was used to assess the palaeogeographic evolution at known opal locations within the Great Artesian Basin, resulting in a set of locality-based descriptors/time-varying patterns. Substantial consistency between the various patterns was identified, which was used as a criterion for constructing a derived set of patterns to be used as prototypes for assessing other unknown localities. Matching the targeted patterns to the greater part of eastern Australia revealed a striking association within the Great Artesian Basin (a renowned area for gem quality opal occurrence), with the results demonstrating that only approximately 11% of this region has a favourable palaeogeographic history, which could be a useful indicator for early stage opal exploration. A data-mining approach for extracting association rules comparing targeted geological locations from arbitrary ones identified a number of geological criteria that are very specific to opal locations, and may thus be used to help understand the opal formation process.

Acknowledgements

This research was funded by ARC Grants FL0992245 and DP0987604. We would also like to thank the reviewers for their constructive feedback.

References

Alley, N., 1998. Cainozoic stratigraphy, palaeoenvironments and geological evolution of the Lake Eyre Basin. Palaeogeography, Palaeoclimatology, Palaeoecology 144, 239–263.

Anand, R., 2001. Evolution, classification and use of ferruginous regolith materials in gold exploration, Yilgarn Craton, Western Australia. Geochemistry: Exploration, Environment, Analysis 1, 221–236.

Barnes, L.C., Townsend, I.J., 1990. Opal deposits in Australia. in: Hughes, F.E. (Ed.), Geology of the Mineral Deposits of Australia and Papua New Guinea, vol. 1, Victoria. The Australian Institute of Mining and Metallurgy.

Boyden, J., Müller, R., Gurnis, M., Torsvik, T., Clark, J., Turner, M., Ivey-Law, H., Watson, R., Cannon, J., 2011. Next-generation plate-tectonic reconstructions using gplates. Geoinformatics: Cyberinfrastructure for the Solid Earth Sciences, 95–114.

Campbell, R.J., Haig, D.W., 1999. Bathymetric change during Early Cretaceous intracratonic marine transgression across the northeastern Eromanga Basin, Australia. Cretaceous Research 20 (4), 403–446.

Carr, S., 1979. Andamooka precious stone fields (map), scale 1:250,000. Geological Survey of South Australia.

Carter, P., 2005. Opal fields lightning ridge region (map), scale 1:100,000. Geological Survey of New South Wales.

Demsar, J., Zupan, B., Leban, G., Curk, T., 2004. Orange: from experimental machine learning to interactive data mining. In: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases.

Exon, N.F., Senior, B.R., 1976. The Cretaceous of the Eromanga and Surat Basins. BMR Journal of Australian Geology and Geophysics 1, 33–50.

Frakes, L.A., Burger, D., Apthorpe, M., Wiseman, J., Dettmann, M., Alley, N., Flint, R., Gravestock, D., Ludbrook, N., Backhouse, J., 1987. Australian Cretaceous shorelines, stage by stage. Palaeogeography, Palaeoclimatology, Palaeoecology 59, 31–48.

Gaillou, E., Delaunay, A., Rondeau, B., Bouhnik-leCoz, M., Fritsch, E., Cornen, G., Monnier, C., 2008. The geochemistry of gem opals as evidence of their origin: Ore geology reviews. Ore Geology Reviews 34, 113–126.

Geological Survey of Queensland, 2012. Interactive Resource and Tenure Maps [Digital Dataset]. Department of Mines and Energy.

Landgrebe, T., Mueller, R., 2008. A spatio-temporal knowledge-discovery platform for Earth-Science data. Image Computing Techniques and Applications (DICTA) (February), 394–399.

Langford, R.P., Wilford, G.E., Truswell, E.M., Isern, A., 1995. Palaeogeographic Atlas of Australia, volume 10, Cainozoic. Australian Geological Survey Organisation, Canberra.

Merdith, A.S., Landgrebe, T.C.W., Dutkiewicz, A., Müller, R.D., in press. Towards a predictive model for opal exploration using a spatio-temporal data mining approach. Australian Journal of Earth Sciences 60, 1–13. http://dx.doi.org/10.1080/08120099.2012.754793.

Qin, X., Müller, R., Cannon, J., Landgrebe, T., Heine, C., Watson, R., Turner, M., 2012. The gplates geological information model and markup language. Geoscientific Instrumentation, Methods and Data Systems 2, 1–63.

Raza, A., Hill, K.C., Korsch, R.J., 2009. Mid-Cretaceous uplift and denudation of the Bowen and Surat Basins, Eastern Australia: relationship to Tasman Sea rifting from apatite fission-track and vitrinite-reflectance data. Australian Journal of Earth Sciences 56, 501–531.

Robertston, R., Scott, D., 1989. Coober Pedy Precious Stone Field (Map): Scale 1:250,000. Department of Mines and Energy, South Australia.

Smallwood, A., 1997. A new era for opal nomenclature. Australian Gemmologist 19 (12).

Tan, P., Steinbach, M., Kumar, V., 2006. Introduction to Data Mining. Addison-Wesley.

Thiry, M., Milnes, A.R., Rayot, V., Simon-Coinçon, R., 2006. Interpretation of palaeoweathering features and successive silicifications in the Tertiary regolith of inland Australia. Journal of the Geological Society 163 (4), 723–736.

Totterdell, J., 2002. Palaegeographic Atlas of Australia (a set of 10 volumes). Geoscience Australia GIS Dataset-National 1–10, November.

Veevers, J., 2000. Analysis of australia: present tectonic state. In: Veevers, J.J. (Ed.), Billion-year Earth History of Australia and Neighbours in Gondwanaland. McPhersons Printing Group, Maryborough.