

DR STEPHEN A. BOWDEN (Orcid ID : 0000-0002-6370-8175)

DR AYAD FAQI (Orcid ID : 0000-0002-9142-1577)

Article type : Original Article

**Modelling the Shimokita deep coalbed biosphere over deep geological time: Starvation, stimulation, material balance and population models**

Stephen A. Bowden<sup>1\*</sup>, Abdalla. Y. Mohamed<sup>1</sup>, Ayad. N. F. Edilbi<sup>1,2</sup>, Yu-Shih Lin<sup>3</sup>,  
Yuki Morono<sup>4,5</sup>, Kai-Uwe Hinrichs<sup>6</sup> and Fumio Inagaki<sup>4,5,7,8</sup>

<sup>1</sup>School of Geosciences, University of Aberdeen, Aberdeen, AB24 3UE, Scotland, UK,

<sup>2</sup>Present address: Department of Petroleum Geosciences, Faculty of Science, Soran University, Soran, Kurdistan Region, Iraq,

<sup>3</sup>Department of Oceanography, National Sun Yat-sen University No. 70, Lien-hai Road., Kaohsiung 80424, Taiwan, R.O.C.

<sup>4</sup>Kochi Institute for Core Sample Research, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Nankoku, Kochi 783-8502, Japan

<sup>5</sup>Research and Development Center for Submarine Recourses, Yokosuka 237-0061, Japan

<sup>6</sup>MARUM-Center for Marine Environmental Sciences, University Bremen, Leobener Straße 8, 28359 Bremen, Germany

<sup>7</sup>Present address: Mantle Drilling Promotion Office, Institute for Marine-Earth Exploration and Engineering, Yokohama 236-0001, Japan

<sup>8</sup>Research and Development Center for Ocean Drilling Science, JAMSTEC, Yokohama, 236-0001 Japan

\*Corresponding Author

Email: s.a.bowden@abdn.ac.uk

Tel: +44(0)1224 273467

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/bre.12399

This article is protected by copyright. All rights reserved.

## Abstract

Basin models can simulate geological, geochemical and geophysical processes and potentially also the deep biosphere, starting from a burial curve, assuming a thermal history and utilising other experimentally obtained data. Here we apply basin modelling techniques to model cell abundances within the deep coalbed biosphere off Shimokita Peninsula, Japan, drilled during Integrated Ocean Drilling Program (IODP) Expedition 337. Two approaches were used to simulate the deep coalbed biosphere: 1) In the first approach, the deep biosphere was modelled using a material balance approach that treats the deep biosphere as a carbon reservoir, in which fluxes are governed by temperature-controlled metabolic processes that retain carbon via cell-growth and cell-repair and pass it back via cell-damaging reactions. 2) In the second approach, the deep biosphere was modelled as a microbial community with a temperature-controlled growth ratio and carrying capacity (a limit on the size of the deep biosphere) modulated by diagenetic-processes. In all cases, the biosphere in the coalbeds and adjacent habitat are best modelled as a carbon-limited community undergoing starvation because labile sedimentary organic matter is no longer present and petroleum generation is yet to occur. This state of starvation was represented by the conversion of organic carbon to authigenic carbonate and the formation of kerogen. The potential for the biosphere to be stimulated by the generation of carbon-dioxide from the coal during its transition from brown to sub-bituminous coal was evaluated and a net thickness of 20 meters of lignite was found sufficient to support an order of magnitude greater number of cells within a low-TOC horizon. By comparison, the stimulation of microbial populations in a coalbed or high-TOC horizon would be harder to detect because the increase in population size would be proportionally very small.

## Introduction

Sedimentary basins are increasingly recognised as microbial habitats spanning long periods of geological time during which individual communities develop that are unique to a formation, or even a lithological component of that formation (Omar, 2003; Breuker et al., 2011; Inagaki et al., 2003; Nunoura et al., 2009; Hoshino et al., 2011). One such example is the Shimokita coalbeds from NE Japan (Figure 1; Inagaki et al., 2015) drilled down to ~2.5 km below the ocean floor during IODP Expedition 337- Deep Coalbed Biosphere off Shimokita. The coalbeds encountered at IODP-337 Site C0020 are overlain by diatomaceous marine sediments in which there is a high abundance of microbial cells ( $10^7$ -  $10^9$  cells.cm<sup>-3</sup>)

suggesting, an active deep biosphere. However, the coalbeds are separated from the overlying marine sediments by sedimentary formations with very low cell-abundances ( $< 10^3$  cells.cm<sup>-3</sup>) effectively making the coalbed an oasis of microbial life. The microbial community has a number of genetic markers that link the coalbed microbiome to an estuarine depositional environment, e.g. the environment in which the coalbeds were originally deposited (Inagaki et al., 2015; Gross et al., 2015). Thus, the implication is that most microbes within the coalbed originated from the formation's depositional environment and have adapted to a new life-habit that proceeds slowly over long geological durations (Inagaki et al., 2003; 2015; Liu et al., 2017; Trembath-Reichert et al., 2017).

Microbial communities within sediments have been accepted to specialise or adapt to changing substrate availability with increasing depth and *in situ* temperature (Emerson et al., 1980; Iversen and Jørgensen, 1985; Horsfield et al., 2006). The Shimokita coalbeds offer another perspective and suggest that over geological time burial within a basin causes microbial communities within a sedimentary horizon to develop as distinct formation-bounded microbiomes that are ultimately the product of long acting and in many cases far-field geological processes. Therefore, aside from a selection based on substrate availability, a microbiome within a formation will also reflect such things as diagenetic and tectonic history.

Modelling microbiomes within sedimentary formations over geological time thus requires a basin model that can convert processes initiating and deriving from subsidence and diagenesis (Mckenzie, 1978; Waples, 1994) into input that can simulate the deep biosphere. To formulate such models, it is necessary to consider what processes are most important for the deep biosphere within sediments and sedimentary rocks, and therefore what might realistically be the minimum needed to account for a population over deep geological time.

#### *Sediment-hosted deep biosphere*

Previous work seeking to understand the biosphere present within sediments has initiated its investigation from shallow to deep locations (e.g., marine sediments and oil fields). Within marine sediments (Morita & ZoBell 1955), the deep biosphere has been shown to persist with depth and burial (Parkes et al., 1994, 2000), logarithmically declining in a manner that permits the prediction of its size in deeper horizons by extrapolation from the seafloor. Within the first few meters, microbial populations undergo a selection as the availability of substrates and other factors change (Amend and Shock, 2001), and in this process a

subsurface biosphere is passed into the deep biosphere and potentially deep geological time. In addition to geochemical changes, there are also physical parameters that change with depth such as temperature, porosity and pressure etc. and all of these at an empirical level can be shown to be limiting factors (Parkes et al., 2000). In the case of Site C0020 explored during Expedition 337, petrophysical characteristics have been notably associated with instances of high and low microbial activity (Tanikawa et al., 2018).

Cell concentrations at depth can be extrapolated from cell concentrations in overlying sediment, providing that there is geological continuity between the two points. This procedure has been used to provide estimates of the deep biosphere's population size (Kallmeyer et al., 2012; Hinrichs and Inagaki, 2012; McMahan and Parnell, 2014; Magnabosco et al., 2018; Hoshino and Inagaki 2019). This perspective originates from studies of Pacific Ocean sediments (Parkes et al., 1994), for which many assumptions about continuity in sedimentation are reasonable. However, ocean sedimentary basins (basins in which the basement comprises oceanic crust) differ substantially to sedimentary basins in continental crust in many respects; basins within oceanic crust are typically younger, do not have as thick a sedimentary fill as basins within continental crust, and in many cases, sedimentation is relatively continuous over long periods (Hay et al., 1998). This is not the case for sedimentary basins overlying continental crust where basins may be much older, subsidence has occurred in multiple phases, sedimentary cover is thick and deposition is often interrupted by uplift and subsidence (McKenzie, 1978). Thus, on long geological time scales, for many sedimentary basins extrapolation across formation boundaries is harder to justify (Meister 2015).

#### *Sedimentary Rock-Hosted Deeper Biospheres*

When petroleum is degraded by microbes at the earth's surface it becomes enriched in asphaltic components (Connan 1984). The inference can thus be drawn that where asphaltic heavy petroleum occurs in subsurface petroleum reservoirs, microbial communities have been active (Head et al., 2003). A potential temperature limit (60-90 °C) is observed for the subsurface occurrence of petroleum that closely corresponds to the pasteurisation temperature of microbes (Bernard and Connan, 1992), generating the idea that oils present in reservoirs beneath this temperature are more likely to be degraded because their microbial communities have not been pasteurised. Within petroleum reservoirs a number of factors have been suggested to limit the deep biosphere including salinity and the diffusion of aqueous

chemical-substrates (Head et al., 2003), the resupply of fresher oil depending on the situation (Horstad and Larter, 1997; Horsfield et al., 2006) and crucially from a basin analysis perspective also burial history (Wilhelms et al., 2001). The recognition of burial history, as opposed to thermal history, is important because of the implied importance of a sequence of geological events (e.g. a cause and effect) as apposed combinations of time and temperature (Waples 1980). This latter aspect is important because combinations of time and temperature, when applied in any sequence, achieve nearly the same eventual level of thermal maturation (Sweeney and Burnham 1990), which is not the case for sterilising processes.

#### *Geological Controls and Kill Mechanisms*

From current literature, lists can be drawn up of sedimentological and geological factors that control the deep biosphere (Parkes et al., 2000), in addition to lists of biological processes (Lever et al 2015). However, the two lists are not linked together by a common modelling approach. This puts studies of the deep biosphere in a similar situation to early studies of mass extinctions, where for example, geological events that may cause or trigger mass extinctions are known, but the mechanism by which this might happen (the kill mechanism) is not specified (Knoll et al., 2007). Because basin models express far-field geological events (e.g. rifting and subsidence) in the form of diagenetic processes (burial metamorphism and compaction) and return physical parameters, they have the potential to link specific sedimentological factors and geological events to specific biological processes that may serve as kill mechanisms in the deep biosphere. We take a kill mechanism, in the context of the deep biosphere, to be a combination of geological, geochemical and biochemical processes that act in combination to reduce cell abundance.

#### *Measures of the deep biosphere*

The main quantitative measure of the sediment-hosted deep biosphere is the number of microbial cells determined per unit volume of sediment (Weinbauer et al., 1998, Parkes et al., 2000; Morono et al., 2009). In this approach the number of cells with DNA stainable by a fluorescent nucleidnucleic acid stain (such as SYBR Green I) within a known volume of sediment are counted. Cell separation and enumeration procedures (Kallmeyer et al., 2008; Morono et al., 2009) in addition to careful cleaning and contamination monitoring (House et al., 2003) are used to increase sensitivity, and also for quality control purposes. If cellular carbon content is known or can be assumed (Romanova and Sazhin, 2010; Braun et al, 2016)

then the microbial population can be expressed as grams of carbon per cubic centimetre of sediment (Kallmeyer et al. 2012).

The deeper petroleum-reservoir hosted biosphere has mainly been measured or quantified by its metabolic products, or by the recalcitrant fraction of petroleum that is undegraded (Peters et al., 2005). However, in many cases, multiple phases of petroleum charging make its measurement by this approach complicated (Horstad and Larter, 1997). However, some metabolic products can be highly specific (Volkman et al., 1983; Jones et al., 2008; Bennett et al., 2007) and in certain cases, where the effects of multiple periods of oil-charging can be compensated for, it has proven possible to model the formation of these products as a temperature-dependent process (Lamorde et al., 2015). As petroleum components are hydrocarbon compounds, their abundance can be expressed as a unit of carbon, permitting rates of turn-over to be calculated (Larter et al., 2003). The sediment-hosted biosphere can also be measured or quantified in similar terms of a rate or turnover (Colwell et al., 2008; Langerhuus et al., 2012; Beulig et al., 2018) and in this case, the unit of measurement would be grams of carbon per unit time. In the following sections, results are expressed as cell abundances, primarily because this provides access to the largest available dataset for model evaluation, but computation is performed in masses of carbon per unit volume of sediment.

#### *Purpose and Scope*

Here are presented two models of the Shimokita coalbed deep biosphere. Both models comprise basin models, but differ in the way in which deep biosphere is modelled. In one model, the deep biosphere is represented as a pool of organic carbon with fluxes in and out driven by metabolic processes that are co-existent with geochemical processes. In the other model, the deep biosphere is considered a population with a temperature dependent growth rate, and carrying capacity limited by geological factors. Finally the capability of basin models to simulate diagenetic processes such as the early-stage generation of carbon dioxide from kerogen is utilized. In this case we calculate how much carbon dioxide might be released during the transition from lignite to sub-bituminous coal, and the potential this process has to stimulate autotrophic methanogenesis within the coalbeds. In this instance we take methanogenesis and carbon dioxide generation within the Shimokita Coalbeds (Inagaki et al., 2015) as one of a number of coupled metabolic and geochemical process that might be suited to simulation by basin modelling.

## **Geology and Basin Model**

### *Assumptions and Workflow*

Basin modelling was performed within Petromod 2015 and 2016. Using data exported from a basin model, models of the biosphere were computed as finite difference models within R. This workflow is summarised in Figure 2 and the formulation of deep-biosphere models is explained later. In all cases it was assumed that carbon (including that within microbial cells), along with other mass were not transferred across Unit-boundaries, an assumption that fits observed fluid-anomalies within mud gases and porewater (Inagaki et al., 2013, 2015).

### *Modelled Units*

The following lithological description is summarised from Inagaki et al., (2013). Site C0020 is located near to the Shimokita Peninsula, Japan, (1180 m water depth) in a forearc basin formed by the subduction of the Pacific plate under the Okhotsk plate (Figure 1). The total subsurface depth modelled is 2500 m and the total depth of the hole is 2466 m subsurface or 3646 meters subsea. Here the formations previously described (Inagaki et al., 2013; Gross et al., 2015) are regrouped into Units for basin modelling purposes, which are described from oldest to youngest (Table 1 and Figure 3). From 2466 to 1000 mbsf the succession comprises sedimentary units deposited in near shore, estuarine or deltaic settings and comprises mudstones, siltstones, sandstones and coal beds. Within the oldest and deepest sections coal horizons are sparse, and based on pollen flora indicate a late Oligocene age (Depositional Unit I), coalbeds, are also present in Unit I. Coal-bearing horizons are more common above this interval and these coalbeds have a middle Miocene age (Depositional Unit II). A higher proportion of coalbeds within Unit II has several implications from a modelling perspective, here picked up upon only slightly, that include such things as mechanical and physical properties; e.g. vertical permeability via the formation of cleats, thermal conductivity, a higher compressibility (Tanikawa et al., 2016). Muddier units, interbedded with occasional sandstones, overly the coal seams and these contain plant macrophytes (Depositional Unit III). The top of interval of the Miocene estuarine facies is siltier and contains fragments of coal (Depositional Unit IV). From the 1000 to 0 m the section comprises poorly consolidated and weakly lithified marine mud, with a base that contains occasional silt and sand. The marine mudstone in Unit V effectively records a marine transgression and the onset of subsidence to present day water depths. The marine mudstone and mud are split into an upper and lower section (Depositional Unit V and Unit OB), to allow for an overburden unit for which data was not as extensively gathered. The lithological composition of each unit is tabulated in



Table 1 along with other geochemical characteristics. A burial history and a lithology column for Site C0020 is shown in Figure 3 and the physical properties used presented in Table 2.

#### *Far Field Aspects; Tectonics, Heat Flow and Other Boundary Conditions*

According to Honda (1985) present day heat flow is about 20 to 50 mW.m<sup>-2</sup> near to the trench and rises to more than 120 mW.m<sup>-2</sup> at the volcanic front. The present heat flow at the Japanese Sea to the west is about 92 mW.m<sup>-2</sup>. This present-day heat flow at C0020 was determined by Tanikawa et al (2016) to be 30 mW.m<sup>-2</sup>, based on Horner-corrected bottom hole temperatures.

The location of Site C0020 relative to these zones of distinctive heat flow has changed over time as the volcanic front backstepped slightly during a period of slab retreat concurrent with the opening of the Japan Sea from 30 Ma to 15 Ma, before returning to its cooler present-day situation (Jolivet et al., 1994 citing Ohguchi et al, 1989). This tectonic change was reflected in the basin model by allowing Site C0020 to cool from 15 Ma onwards to its present day heat flow (representing the westward movement of the loci of volcanic activity). The very high levels of heat flow and the corresponding changes in coal rank seen in Northern Japan (Honda 1985) are not evidenced at Site C0020, and to yield models that could match thermal maturity calibration data (described later), it was found necessary to keep heat flow in the 30-20 Ma period to less than 60 mW.m<sup>-2</sup>.

Sea level was taken to have changed from a stormbase maximum depth of 60 m during the deposition of the estuarine facies to the present-day value of 1180 m. Sediment seawater interface temperatures were obtained using assumptions based on palaeolatitude (Hantschel and Kauerauf, 2009). In the case of Site C0020, sediment water interface temperatures mainly vary as a function of water depth, rather than latitude, and there is change in sediment water interface temperature of 20 °C to 5 °C that initiates in the mid-Miocene.

#### *Temperature and Thermal maturity*

The highest, Horner-corrected, bottom hole temperature is 63 °C and this along with other indicators demonstrates that the modelled interval is thermally immature with respect to petroleum generation (Inagaki et al., 2015). The methane present is depleted with a  $\delta^{13}\text{C}$  of -65 ‰ and contains a low proportion of larger hydrocarbons (C1/C2 hydrocarbon gases > 100). The vitrinite reflectance values are also low, generally less than 0.4 % (Inagaki et al., 2013), and organic macerals within the younger Miocene coalbeds contain huminite,



indicating that vitrinite formation is not complete (Gross et al., 2015).  $T_{\max}$  values indicate pre-oil window if not pre-vitrinite reflectance values in most units (Table 1), and pre-oil window biomarkers such as sterenes and hopenes are dominant confirming that this is the case for most of the interval, with the onset of hopane-isomerisation (percentage of  $C_{32}$   $14\alpha,17\beta(H)$  22 S hopane) beginning to be measureable at a depth of ~3000 meters subsea (Figure 3). While superficially this makes the Shimokita coalbeds appear intermediate between the “deep” and “deeper” biosphere models described earlier, on the other hand it completely changes the availability of organic substrates. This is because the most labile sedimentary organic matter has been remineralised (Westrich and Berner, 1984), leaving behind refractory organic matter and discrete organic molecules resembling the pre-oil window type of biomarker that contain either functional groups or points of unsaturation that permit the binding of organic molecules into kerogen (Kohnen et al., 1991). Thus kerogen within Units I-IV is not a potential source of lower molecular weight substrates, instead lower molecular weight substrates are bound into kerogen or oxidised and their carbon sequestered in carbonate. These aspects may thus be important and were incorporated into models of the biosphere described later.

Calibration using thermal maturity parameters for organic matter within the Miocene coalbeds is complicated – vitrinite reflectance, in this case, is of limited use as the formation of vitrinite is not complete and thus the parameter varies little (Figure 3). For Units I and II (Miocene and younger) the % 22S  $C_{32}$  hopane parameter was used (Sajgó and Lefler, 1986). This inflects measurably at the base of the hole when measured in core-samples of coal or shale (Figure 3d), although in the shallower units, where drill-cuttings form the main type of sample, the parameter would not be predicted to vary and the observed variation is due to reworked organic matter (re-sedimented dinocysts were observed; Inagaki et al., 2013). Here a constant heat flow model (a model in which the heat flow is set at an average but constant value that still reasonably predicts present day heat flow and thermal maturation shown in Figure 3b), under predicts thermal maturation as it does not take account of hotter temperatures in the geological past (Figure 3d).

In Unit III and IV a model of the Raman R1 parameter was used (Muirhead et al., 2012). This parameter is heating rate dependent and is only applicable to very low levels of thermal maturation, or heating over geologically short durations. At metamorphic levels of thermal alteration, a simply measured R1 parameter (e.g. obtained by fitting curves to represent the G

and D peaks), decreases with increasing thermal alteration (Wopenka and Pasteris 1993). However, for high heating rates (Zhou et al., 2014) or pre-oil window thermal alteration (Muirhead et al., 2012), the parameter rises before falling. The value is over-predicted by a constant heat flow model for Units III and IV, which were deposited subsequent to the onset of cooling and subsidence in the Pliocene (Figure 3).

## **Biosphere Modelling**

### *Material Balance Model*

A schematic representation of a material balance model of the deep biosphere is shown in Figure 2b and a list of terms is provided in Table 3. The model comprises a box for carbon that is metabolically active and part of a cell within the deep biosphere and a box for the total organic carbon present in sediment, but that is not within the deep biosphere. This is effectively a reconceptualisation of the “carbon-sink” biospheres described by Lipp et al. (2008) and Kallmeyer et al. (2012), and the Price and Sowers (2004) concept of a subsurface biosphere controlled by fundamental metabolic processes. The total carbon within the sediment is then subdivided into four further fractions; 1) carbon bioavailable as dissolved organic carbon (DOC), 2) organic carbon that is not bioavailable but neither is it bound within kerogen and later referred to as  $O_{Cr}$ , 3) carbon bound within kerogen, and 4) carbon removed from bio-availability by its sequestration as carbonate mineral phases. Growth or cell-maintenance incorporates organic carbon from the pool of total organic carbon (TOC) within sediments, but only via the fraction of dissolved organic carbon, whereas cell-damaging reactions remove organic carbon from cells living in the deep biosphere. This model is clearly not physiologically representative of a cell because it is too simple, and nor is it a model of a community described later, but rather an approximation of Earth or geological systems where biogeochemical reactions play a key mediating role. The main difference between the material balance and population models described later is that the potential limits of the size of the deep biosphere is not explicitly specified in the material balance model.

Changes in the amount of organic matter in the deep biosphere within the simple system shown in Figure 2b can be formulated with the following finite difference calculation:

$$N_i = N_o + N_g - N_d$$

Equation 1

Where  $N_i$  and  $N_o$  are the mass of organic carbon within the deep biosphere at two points in time,  $N_g$  is mass of organic carbon transferred to the biosphere by biological processes and  $N_d$  is the loss of organic carbon from the deep biosphere as a consequence of damaging reactions. The units of measurement for all of these variables is  $\text{gC.cm}^{-3}$  of sedimentary rock or sediment.

The loss of organic carbon from the deep biosphere by a single process ( $N_d$ ) is represented by the following:

$$N_d = F_{dam} \times [T \times k_{dam} \times t] \times N_o$$

Equation 2

Where  $t$  is the duration, in a finite difference model this will be the length of a time step,  $k_{dam}$  the temperature dependent rate for a cell-damaging, biomolecule-damaging or biomolecule-altering process, and  $T$  the temperature at which the reaction takes place, for a finite difference model this will be the temperature at the mid-point of a single timestep.

Temperature in this case refers to palaeotemperature over geological time and is obtained from a basin model, with only the present-day temperature being a direct measurement. To create an empirical model that fits observations any number of processes can be aggregated, and their effects weighted via  $F_{dam}$ , a weighting factor. For the models presented later depurination and racemisation (two protein damaging reactions) were considered (Price and Sowers 2004; Lever et al. 2015); although clearly a far broader spectrum of processes would ideally be included.

The flux of organic carbon into the deep biosphere ( $N_g$ ) is formulated as follows:

$$N_g = F_{rep} \times [T \times k_{rep} \times t] \times [D_{ac} \times O_{Cr}]$$

Equation 3

Where  $T$  and  $t$  represent temperature and time as described above for equation 2 and  $k_{rep}$  is a temperature dependent biological process and  $F_{rep}$  its weighting factor; in this case the median rates of repair and growth given for methanogens presented in Price and Sowers (2004).  $O_{Cr}$  is the carbon not bound in kerogen or sequestered in carbonate, only a fraction of which ( $D_{aC}$ ) is available to the deep biosphere (labelled 2) in Figure 2b).  $O_{Cr}$  is high early during early diagenesis but will decline as sediment is turned into sedimentary rock and lithifies. The  $D_{aC}$  and  $O_{Cr}$  relationship is one of several relationships which were obtained empirically, and these are described later (Equation 8).  $O_{Cr}$  and  $N_g$  have the units  $\text{gC.cm}^{-3}$  and  $F_{rep}$  and  $D_{aC}$  are factors without units.

During the early stages of burial sediment contains high proportions of labile organic carbon and at this stage the  $O_{Cr}$  parameter will be high (Westrich and Berner 1984). But this situation does not exist indefinitely as initially labile organic carbon becomes refractory (and is no longer bioavailable) and a sediment is converted to sedimentary rock. This process is captured in Equation 4 which determines  $O_{Cr}$  as function of TOC, density and porosity (describing the conversion of a sediment to rock) and the loss of carbon to early stage diagenetic processes.

$$O_{Cr} = TOC/100 \times \phi/100 \times \rho \times C_{loss}$$

Equation 4

$TOC$  is the total organic carbon content of a formation and obtained by measurement;  $\phi$  its porosity and  $\rho$  the density. ( $TOC$  and porosity are here divided by 100 to convert from percentages, density has the units  $\text{g.cm}^{-3}$ .) The present day values of porosity and density are obtained by measurement, but it is necessary to know porosity and density at specific points in geological time. Therefore their past-values are obtained from a basin model, in this case using Athy's law (Athy, 1930), with only the present-day value being a direct measurement, and this itself is an average value representing a formation rather than a single lithology (Table 1).  $C_{loss}$  is the fraction of total organic carbon that is no-longer available because it has been sequestered in kerogen ( $C_{lker}$ ) or carbonate ( $C_{lcarb}$ ), and it is obtained using equation 5.

$$C_{loss} = C_{lker} + C_{lcarb}$$

Equation 5

These carbon-loss parameters ( $C_{lker}$  and  $C_{lcarb}$ ) are transient and change over time during burial, and are calculated from other parameters that themselves change during burial and diagenesis. These burial-dependent processes are given in Equations 6, 7 and 8. As presented, Equations 6, 7, and 8 are empirical relationships obtained by fitting regression models to pre-existing data sets.

The fraction of organic carbon sequestered in kerogen ( $C_{lker}$ ) was computed using Equation 6:

$$C_{lker} = a_{lker} \times V_{Rcalc}^2 + b_{lker} \times V_{Rcalc} + c_{lker}$$

Equation 6

Equation 6 was obtained by fitting a polynomial equation to normalised sterene concentrations (e.g.  $Max_{sterene} - X_{sterene} / Max_{sterene} - Min_{sterene}$ ) and vitrinite reflectance, yielding the exponents  $a_{lker}$ ,  $b_{lker}$  and  $c_{lker}$ . The rationale for using sterene concentrations is explained in the following paragraph. The values of exponents found for C0020 are shown in Table 4.  $V_{Rcalc}$  is vitrinite reflectance at a given point in geological time and its units are percentage light reflected. Vitrinite reflectance is measured for the present-day sediments, but in this case it also refers to a specific point in geological time (the past) and must be obtained from a basin model. In this study, this was done within Petromod using the EasyVit model (Burnham and Sweeny, 1989; Sweeney and Burnham, 1990).

The Shimokita coal beds are pre-oilwindow with respect to oil and gas generation, and their organic macerals still contain huminite (Gross et al., 2015), indicating that vitrinite and kerogen are still forming. During the early stages of burial diageneses, kerogen sequesters organic carbon by forming geomacromolecules from more labile components (Tissot and Welte 1984). Sterenes are one of many labile components of sedimentary organic matter and contain a site of unsaturation by which they are bound within kerogen during early diagenesis (Kohnen et al., 1991). Sterene concentrations generally decrease downhole at Site C0020 (Figure 5a – data from Inagaki et al., 2013), and their decreasing concentration is here taken as a measurable proxy of labile organic carbon within kerogen. That is not to say that either sterenes or huminite are major biosphere substrates, or that sterenes are bound within huminite (fossil wood), but that the processes co-occur and are used in the same way as petroleum biomarker thermal maturity parameters, that use processes such as isomerisation to proxy for catagenesis and oil and gas generation (Peters et al., 2005).

$C_{lcarb}$  is the fraction of total carbon sequestered by the formation of diagenetic carbonate and was calculated using Equation 7:

$$C_{lcarb} = m_{lcarb} \times \log \phi + c_{lcarb}$$

Equation 7

Equation 7 was obtained by fitting a straight line to inorganic carbon concentrations expressed as the fraction of total carbon (e.g. total inorganic carbon/total carbon) and porosity (Figure 4b), yielding the exponents  $m_{lcarb}$  and  $c_{lcarb}$ . The values of these exponents for Site C0020 are shown in Table 4. To obtain this empirical relationship data were taken from the estuarine facies (Units III and IV) and any anomalously high values of TIC were ignored (e.g. samples containing shell fragments etc.). The estuarine facies alone was chosen because it has very little detrital carbonate relative to the marine facies (therefore changes in inorganic carbon better represent authigenic as opposed to sedimentary processes). Carbonate concretions and authigenic carbonate cements form during early diagenesis as distinct authigenic phases and pore filling cements (Berner, 1968; Wilkinson and Damper, 1990). At Site C0020 carbonate concretions are found throughout the drilled-interval but are notable in freshwater parts of the estuarine facies, where stable isotope data show that they are closely tied to processes such as methanogenesis (Phillips et al., 2018). The calculation of porosity over geological time is described above for Equation 4, and was done using Athy's law (Athy, 1930). As for the calculation of changes in sterene concentration as a function of vitrinite reflectance, there need not be a direct mechanistic link between the two parameters, but instead increased carbonate-cementation is taken as being contemporaneous with porosity reduction during burial.

$D_{ac}$  is the fraction of  $O_{Cr}$  or labile sedimentary organic carbon available to the deep biosphere and was calculated using Equation 8:

$$\log(D_{ac}) = m_{DaC} \times \log TOC + c_{DaC}$$

Equation 8

Equation 8 was obtained by fitting a straight line (Figure 4c) to TOC and dissolved organic carbon concentrations for shallow sediments from the Weddell Sea (Egeberg and Abdullah 1990). The value of the exponent  $m_{D_{ac}}$  is shown in Table 4, while the  $c_{D_{ac}}$  parameter is sufficiently close to zero that it was left out of calculations. The Egeberg and Abdullah (1990) dataset for the Weddell Sea its TOC values are similarly low and close to those of the Shimokita-sandstones and -siltstones within the estuarine facies (Table 2). The dataset from Egeberg and Abdullah (1990) also concerns sediments that have not been buried deeply and have a mild thermal history, thus much of the TOC is still labile and thereby provides a better estimate of  $D_{ac}$  as function of  $O_{Cr}$ , as opposed to  $D_{ac}$  as a function of a mixture of refractory and labile organic carbon.

Figure 5 illustrates downhole variation in sterene and total inorganic carbon and associated calculations of carbon loss to the formation of kerogen and authigenic carbonate. At greater depths generally a higher fraction of carbon is lost to kerogen (Figure 5b), whereas the varied lithology of the main coalbed horizon leads to high rates of authigenic carbonate formation – particularly in the coal and sandstone rich formations (Figure 5d).

#### *Population Model*

A schematic representation of a deep-biosphere population model is shown in Figure 2c. The model comprises a deep-biosphere population, the potential size of which is limited by the carrying capacity of the host formation. As the capability of the population to grow to fill its carrying capacity changes, or as the carrying capacity of the formation adjusts because of changes in the form and type of organic carbon, the size of the population adjusts accordingly. The carrying capacity of any formation was taken as the organic content available to the biosphere and the organic carbon content initially present within cells.

Carrying capacity was determined using:

$$cc = (D_{ac} \times O_{Cr}) + N_0$$

Equation 9

Where  $D_{ac}$  and  $O_{Cr}$  are as defined above for equations 3 and 4 , and  $N_0$  an initial concentration of cells expressed in  $\text{gC.cm}^{-3}$ .



The deep-biosphere population model presented here is a negative-reconceptualisation of the deep biosphere presented in Horsfield et al. (2006); it is negative because in this case diagenetic processes occurring during burial will convert substrates to non-bioavailable forms (kerogen and carbonate), as opposed to generating substrates from kerogen.

The Population Model formulated for a finite difference calculation is shown below:

$$N_i = N_o + a \times t \times N_o \times ((cc - N_o) / cc)$$

Equation 10

Where  $N_i$  and  $N_o$  are the final and initial concentration of cells at two points in time, as described above for equation 1. Computationally these numbers are handled as a mass of organic carbon. The growth ratio  $a$ , represents the change in gC per unit time,  $cc$  is the carrying capacity. The growth ratio is temperature dependent, and within a finite difference model this is the temperature at the mid-point of a time step. As temperature within a sedimentary formation undergoing burial varies over geological time, temperature is determined using a basin model. To keep parity with the material balance model described above the growth ratio  $a$  was determined as function of cell-damaging and cell-repairing reactions;

$$a = (T \times k_{rep}) / (T \times k_{dam})$$

Equation 11

where  $T$ ,  $k_{rep}$  and  $k_{dam}$  are as defined for equations 3 and 4.

### *Comparison of Deep Biosphere Models*

To illustrate differences in the behaviour of the population- and material balance-biosphere models the effects of growth ratios of 0.9 and 1.1 were considered. For these models the labile organic carbon available to the deep biosphere ( $O_{Cr}$ ) was initially decreased over a short timespan to perturb the biosphere and induce rapid changes, but thereafter the  $O_{Cr}$  parameter was reduced slowly (Figure 6). This condition for carbon availability represents that commonly held to apply to sedimentary organic carbon - where labile and refractory

Accepted Article

pools are observed and the labile organic carbon decays rapidly, while a second pool of sedimentary organic carbon decays slowly (Westrich and Berner 1984). Under these conditions, when the growth ratio is less than 1, and if only the endpoints are considered the two models of the deep biosphere reach similar values (Figure 6a). Thus were the population- and material balance-biosphere models being applied in a situation where sedimentary organic matter was slowly being degraded, and in an environment unfavourable to growth, they would be expected to predict similar end-values.

However, there are differences and most notable of these is that the population model responds to early short-term perturbation but the material balance model does not (Figure 6b and 6c). A further differentiation between the two models occurs when the growth ratio is greater than 1, but the carbon supply is still limited. Under these conditions the material balance model predicts growth, but the deep biosphere model does not (Figure 6c). The primary cause of this different behaviour is the mechanistic action of the carrying capacity term, which acts as a limit within the population model. The population model predicts that an underpopulated biosphere will raise to its limit should the growth ratio permit this, but an overpopulated biosphere will rapidly decline, and subsequently recover upto the limit provided by the carrying capacity. The material balance model does not have this mechanism. Thus both types of biosphere model might reasonably be applied to the Shimokita coalbeds and be expected to provide similar but not identical results. However, if erroneous assumptions are made about growth or carbon supply, then the two models should significantly diverge, and a material balance model would uptake unrealistic quantities of organic carbon. Therefore, in combination, the two models provide a falsifiability that permits erroneous assumptions about either growth within the deep biosphere or carbon cycling within the subsurface to be identified.

## Results and Discussion

### *Results –Material Balance*

Basin model predictions of the Shimokita deep biosphere are shown in Figure 7 for different combinations of basin and biosphere models. The cases are:

- Case 1: variable heat flow basin model & material balance biosphere model (Figure 7a),
- Case 3: variable heat flow basin model and population biosphere model (Figure 7b).
- Case 2: constant heat flow basin model and material balance model (Figure 7c),

In each case predictions are made for the lowest and highest TOC values within each Unit (Table 1). Burial curves for selected models are shown in Figure 8.

All cases exhibited a drop in cell concentrations from  $\sim 10^8$  cells.cm<sup>-3</sup> at the top of the marine mudstones to  $10^4$  cells.cm<sup>-3</sup> at the base of the hole (Figure 7a-c). The four-order magnitude of decline in cells within the marine mudstones (Units OB and V) is consistent with previous work on ocean sediments, and also for Site C0020, and could be predicted by a simple regression based on depth C0020 (Parkes et al., 2000; Inagaki et al., 2015). However, an important feature captured by the combined biosphere and basin models is the co-occurrence of high and low cell abundances within Estuarine Units II to IV, often at nearly equivalent depths. Such features could not be predicted by regression analysis or application of the Parkes depth-regression approach as similar depths would predict similar abundances of cells. However, Cases 1 and 2 (Figure 7a & b) that combine a basin model with a variable heat flow and a material balance biosphere model, and take account of both high and low TOC contents predict a range in cell-concentrations. This is most evident in Units II and III, where higher abundances of cells are predicted for TOC-rich lithologies and lower cell abundances are predicted for the TOC-poor lithologies. This range is not evident for predictions based on a constant heatflow model, that does not take account of changes in thermal environment over time (Figure 7c). Thus it can be concluded that models with thermal histories representing a declining heat flow at Site C0020 better capture the contrast between cells present in the nurturing habitat offered by the coal beds, and cells present in low TOC lithologies. Examining the burial history of units it can be shown that significant declines in cell-concentration occur when the high TOC units are buried to temperatures exceeding

30 °C, and when TOC-poor units are buried to temperatures exceeding 20 °C (Figure 8a and 8b).

Aside from demonstrating that combined basin and biosphere models can predict cell concentrations, it is also necessary to demonstrate that the approach is falsifiable, e.g. bad basin models yield bad predictions of the deep biosphere. Cell concentrations in low-TOC lithologies in Unit III are over-predicted by basin models with a geologically inaccurate constant heat flow (Figure 7). This is because the constant heat flow basin model defers (delays) exposures to temperatures exceeding 20 °C (temp at which the material balance biosphere model causes starvation in low-TOC units), and the onset of temperatures exceeding mesophilic conditions occurs later in the model's burial history thereby permitting high cell concentrations to persist (Figure 8 c & d).

#### *Kill Mechanisms – Material Balance*

Here we investigate the capability of geological, geochemical and biochemical processes to act as kill mechanisms for the Shimokita Biosphere. This is done by observing the effects of the processes and the ensuing changes in cell concentrations as functions of geological time.

Figures 9a-c illustrate a kill-mechanism, primarily utilising cellular damage (depurination), acting on a low-TOC lithology within Unit III, the upper part of the estuarine facies. In this case depurination is chosen because it is one of the cell-damaging reactions coded within the model, and its increase in rate corresponds to a decline in cell abundance. The biosphere in this case is modelled using material balance. Considering a geological history that includes the end of rifting in the Japan Sea and subsidence of the Shimokita area (the variable heat flow model; see Figure 3b), it can be shown that a decline in cell abundances (Figure 9a) occurred when the rate of depurination increased at about 4 Ma and when formation temperature within the unit rose to greater than 25 °C (Figure 9b). In the constant heat flow basin model this decline in cell abundance does not happen until much later (Figure 8a). For both of these the TOC of the formation, its present-day temperature and initial cell concentration are the same. Differences between the two geological scenarios are specifically the result of varying exposure to cellular damage (Price and Sowers, 2004). The predictions made by the variable heat flow basin model represent far field tectonic processes that buried the formation to a depth at which it was exposed to higher temperatures ~20 °C (Jolivet et al., 1994) and increased crustal heat flow.

Figure 7d-f illustrates a starvation-based kill-mechanism at work within Unit II, the Miocene coalbed horizon. Here the comparison is between high TOC coals and low TOC-sandstone lithologies within Unit II. The lithologies interbed within the same unit and thus they share present day- and past-temperatures. Initially there is little difference in cell concentrations, but differences begin to develop after the formations pass the 30 °C threshold at about 4 Ma (Figure 7d). This corresponds to burial-compaction during subsidence and the limited onset of the formation of vitrinite from huminite (Gross et al., 2015) as shown in Figure 7e.

Mechanistically, this stresses the deep biosphere by making the organic carbon needed cell repair or maintenance unavailable; e.g., the biosphere is starved – compare the black and red lines in Figure 7f denoting the carbon needed and that which is available for growth and maintenance. Starvation occurs because diagenesis has sequestered organic carbon within kerogen and carbonate. Prior to 4 Ma, even though the sandstones contain low TOC, there is still sufficient carbon to maintain the deep biosphere, and only rapid burial in the Pliocene, due to flexure of the Okhotsk Plate during subduction (Jolivet et al., 1994), caused differentiation of cell abundances within the two units.

#### *Results and Kill Mechanisms–Population Model*

Figure 7e illustrates predictions of cell abundances made by the variable heat flow Basin Model and the Population Deep Biosphere Model. Substantial differences in cell abundance are both observed and predicted for the maximum and minimum TOC lithologies in Units I to IV. Figure 8 illustrates cell abundances within Unit IV over geological time for the youngest and siltiest member of the estuarine facies. The initial and catastrophic decline in cell abundances observed for the low TOC lithology in Unit IV (Figure 10a) is caused by the carrying capacity and specifically the TOC-content being insufficient to support the population (Figure 10b); other factors such as rates of cell damage are equal between the two models. Once this adjustment occurs the microbial population subsequently declines relatively little, at least compared to the decline observed for the high-TOC formation which continues to decline. From 7.5 Ma to 7 Ma, a decline in cell abundances in the high TOC lithologies is due to reduced carrying capacity – e.g. substrate availability (Figure 10b).

### *Stimulation of deep biosphere by carbon dioxide generation*

The Miocene coalbeds are not sufficiently thermally mature to generate methane via thermogenic reactions (Inagaki et al., 2015). However, carbon dioxide is generated from coals during early catagenesis at sub-pyrolysis temperatures, and particularly so from humic materials with a limited hydrogen content (van Krevelen, 1950; Braun and Burnham, 1987, 1988).

Carbon dioxide (along with the presence of hydrogen) is one of the primary chemical-substrates capable of supporting methanogenesis within sediments (Zeikus 1977), the others being formate, methanol, carbon monoxide, methylamines and acetate (Daniels et al., 1983). The formation of carbon dioxide from coal is important because methanogens have been shown to utilise carbon dioxide within sediments even in the presence of other substrates (Crill and Martens 1986).

Figure 11a is a downhole plot of the S3 pyrolysis-parameter; a measure of the yield of combustible or oxidisable organic carbon from sedimentary organic matter and thus the capability of sediment to provide carbon dioxide (Tissot and Welte, 1984). The highest potential to provide oxidised organic carbon substrates is found within the Miocene coalbed horizon (Unit II), with some potential also found at the top of the estuarine facies where there are plant macrophytes and detrital-clasts of coal (Unit IV).

Figure 12 illustrates the response of coal and sandstone lithologies to carbon dioxide generation within the Shimokita coalbeds. Within a sandstone lithology, without the provision of carbon dioxide cell abundances decline, but they remain unaffected within a coalbed. Within a sandstone supplied with carbon dioxide during the early Pliocene they are initially sustained, and then rise slightly during the Quaternary. Figure 12a was calculated assuming a type III coal with the reaction kinetics reported in Braun and Burnham, (1987) and Burnham and Braun (1990), a 20 m net thickness of coal seams the variable heat flow model (Figure 3b).

The resulting stimulation of cell abundances from 2, 20 and 40 meters net thickness of coal is shown in Figure 11b and 12b. The quantity of carbon dioxide released becomes significant during the Pliocene (temperature of 30 °C, Figure 12c), and this would start to measurably impact the microbial population of a low TOC sandstone at about 3 Ma (Figure 9a) when

temperatures reach 40°C. Assuming that all the carbon released in this way is available to the deep biosphere a coal seam 2 m thick would even be capable of generating sufficient carbon dioxide to raise cell-concentrations by an order of magnitude within a TOC-starved lithology. In the case of a high-TOC lithology, there is a calculable rise, but this might not be sufficient to be measured, e.g. a change 3E4 to 6E4 cells.cm<sup>-3</sup>. This estimate of the extent of stimulation does not take account of actual speciation in terms of both physical chemistry (e.g., adsorbed or free) or chemical speciation (for example partially oxidised carbon such as volatile fatty acids or phenolic compounds, Glombitza et al. 2016), nor does it realistically deal with mass transfer or diffusion. Thus the cases presented in Figure 12a are a higher bound, but were the currently known ~20 meter thickness of coal to generate carbon dioxide and only 10% of this were bioavailable (so that the effects approximated only the 2 m net-thickness case), this would still be sufficient to measurably sustain a community in a low-TOC lithology (Figures 12b-c).

### *Discussion*

From a basin modelling perspective, the Shimokita coalbeds encountered at Site C0020 are relatively simple in the respect that there is little evidence that substantial fluid generation and migration is taking place. This can be seen in the lack of significant over pressure, the absence of petroleum generation, and concentration-profiles for dissolved species that break at Unit-boundaries (Inagaki et al., 2013;2015). Were it necessary to incorporate such elements then commercially available basin models could reasonably do this (Waples, 1994).

The main capability currently lacking from basin models are kinetic models of organic geochemical parameters able to operate at pre-oil window and at pre-vitrinite levels of diagenetic alteration and geological time scale (Farrimond et al., 1998 and Rabinowitz, 2017 for examples of thermal maturation under pre-oil window conditions). What is particularly important is the capability to model the conversion of huminite to vitrinite as this is likely an important diagenetic zone for the deep biosphere within the Shimokita coalbeds (Gross et al., 2015). Because this modelling cannot be done, there is a limited capacity to measure and then model proxies of thermal maturation, independently of the calculation of the release or sequestration of volatile organic compounds. The temperature-dependent release of such substrates has certainly been measured for both labile (Roussel et al., 2015) and refractory components (Tegelaar and Noble, 1994), but the former is not linked to independent geochemical parameters with kinetic models. Because so many basin modelling variables



interact, the independence of the different stages of calculation (thermal models, compaction, substrate generation etc.) and their calibration is necessary to help qualify why one of many non-unique basin modelling solutions is better than another (Peters and Nelson, 2012).

Organic geochemical parameters that have kinetic models provide a means to do this.

There is additional development that would greatly improve the capability of the biosphere models applied to the Shimokita coalbeds. Foremost of these is the capability to track different and varied metabolic reactions; both the number and type of cell damaging reactions, as well as greater specificity for repair processes. The applicability of a given set of metabolic pathways or processes would ideally be conditioned by either genomic or metabolic measurements (Webster et al., 2006; Hoshino and Inagaki, 2019). Herein, only methanogenesis is developed from a modelling perspective (Price and Sowers, 2004), and though many cell-damaging reactions would be expected to be common, the growth rate and energetic-aspects of other metabolisms may not be shared (Lever, 2012, 2015). Notable diagenetic processes for which there is a strong microbial component identified at Site C0020 include iron reduction, sulphur reduction as well as carbonate formation (Phillips et al., 2017, 2018). Similar distinctions could also be made for different types of organic matter (Middelberg 1989), for example necromass (organic matter from recently dead organisms), could be tracked separately from sedimentary organic matter with different levels of recalcitrance (Lomstein et al 2012).

To incorporate such detail within the material balance model this would make use of weighting factors to represent community structure, much as is done for mixtures of varied kerogen types (Tegelaar and Noble, 1994). In the case of the Population model, this would be achieved using the Lotka-Volterra equations which while commonly termed the “predator-prey” equations can represent both competing and dependent life habits (Wangersky, 1978; Renshaw 1991); e.g., fermentation and methanogenic metabolisms.

## **Conclusions**

Basin models can be used to predict subsurface cell concentrations within sedimentary rocks as a function of geological history. Basin models can be combined with material balance- and population-models of the deep biosphere and model deep biospheres within low and high TOC lithologies. In the basin models presented here, a starvation state was induced by allowing diagenetic processes to sequester carbon, and cases were also evaluated where carbon dioxide generated during diagenesis was shown to be numerically important and

potentially significant sources of carbon. A unique strength of this approach is the ability to link basin processes and far field geological events to diagenetic environment and deep biosphere processes.

## Methods

Data sources: The data in Table 1 and used Figure 3 (except for the Raman data) were obtained from data within in Inagaki et al., 2013. Measured cell concentrations were taken from Inagaki et al., 2015. Other data was taken from the library within Petromod 2016, yielding the average values for the Units shown in Table 2.

Example scripts of both Material Balance and Population Biosphere models written in R and are available in supplementary information. An example input file is provided also. R is free software and is available from <https://www.r-project.org/> (R Core-Team 2014).

Soxhlet extraction and GC-MS analysis of extracts of drill cuttings in Units IV to I was carried out at the University of Aberdeen. Prior to analysis extraneous materials were removed from samples. Soxhlet extraction was performed for 48 hrs using a 93:7 v/v mixture of dichloromethane and methanol. Analysis of whole extracts was performed with an Agilent Technologies (AT) 6890N Network GC (pulsed splitless) system fitted with a 30 m × 250.0 µm i.d film thickness 0.25 µm fused silica DB-5 column coupled to an AT 5975 quadrupole mass selector detector (electron energy 70 eV, source 250 °C) with He as a carrier gas. The GC temperature program for saturated hydrocarbons was hold at 60 °C for 2 min, rising at 20 °C min<sup>-1</sup> to 120 °C, and then rising at 4 °C min<sup>-1</sup> to 290 °C and then holding for 23 min. Analysis was performed using a SIM program with 191 *m/z* ion and a dwell time of 40 ms.

Raman spectra were obtained using a Renishaw inVia reflex Raman spectrometer at the University of Aberdeen. An Ar<sup>+</sup> green laser (wavelength 514.5 nm) was focused on the sample using a DMLM Leica microscope using a 50 × objective. Laser spot size was approximately 1–2 µm diameter and laser power was 50 % (<13 mW delivered to the surface). Acquisition parameters were; 20 spectra accumulated over 20 seconds within the Raman shift range of 500 to 2000 cm<sup>-1</sup>. To characterize spectra the following procedure was applied: 1) cropping of data to the Raman shift region of interest, 2) subtraction of a linear baseline fitted to intersect the minimum values (troughs), 3) fitting of two peaks, whose

properties were never more than 90 % from Gaussian and centred about Raman shifts of 1350  $\pm 50$   $\text{cm}^{-1}$  for the D peak and 1585  $\pm 20$   $\text{cm}^{-1}$  for the O peak.

### **Acknowledgements**

The authors are grateful to all crews, drilling team members, lab technicians and scientists on the drilling vessel *Chikyu* for supporting core sampling and onboard measurements during the *Chikyu* shakedown cruise CK06-06 and the Integrated Ocean Drilling Program (IODP) Expedition 337. This work was supported in part by the Japan Society for the Promotion of Science (JSPS) Strategic Fund for Strengthening Leading-Edge Research and Development (to F.I. and JAMSTEC), the JSPS Funding Program for Next Generation World-Leading Researchers (NEXT Program, no. GR102 to F.I.). All shipboard and shore-based data presented in this manuscript are archived and publicly available on-line in either the IODP Expedition 337 Proceedings through the J-CORES (<http://sio7.jamstec.go.jp/j-cores.data/337/C0020A/>), the PANGAEA database ([www.pangaea.de](http://www.pangaea.de), [doi.org/10.1594/PANGAEA.845984](https://doi.org/10.1594/PANGAEA.845984)), or Inagaki et al., 2015, respectively. Petromod Basin Modelling software was provided by Schlumberger to the University of Aberdeen. This is a contribution to the Deep Carbon Observatory (DCO). SAB wishes to thank HSB for support preparing the manuscript.

### **Data availability statement**

All shipboard and shore-based data presented in this manuscript are archived and publicly available on-line in either the IODP Expedition 337 Proceedings through the J-CORES (<http://sio7.jamstec.go.jp/j-cores.data/337/C0020A/>), the PANGAEA database ([www.pangaea.de](http://www.pangaea.de), [doi.org/10.1594/PANGAEA.845984](https://doi.org/10.1594/PANGAEA.845984)), or Inagaki et al., 2015, respectively.

## Figure Captions

Figure 1. Location map of IODP 337 Site C0020.

Figure 2. a) Graphical illustration of the work flow for modelling the Shimokita deep coalbed biosphere over deep geological time. Graphical illustrations of the b) Material Balance and c) Population Models of the Deep Biosphere.

Figure 3. a) Burial Curve and b), Heat Flow Models and Sea Level Curves illustrating the geological history of C0020. c) Lithological Column and Modelled Depositional Units. d) modelled vs predicted downhole temperatures. e) Modelled and predicted values of the %  $C_{32} 17\alpha, 21\beta(H) 22S/22 S +22R$  parameter. Data shown for core (filled circles) and drill cuttings (open circles) and for predictions made by variable and constant heat flow models (solid and dashed lines). f) Modelled and predicted values of the Raman R1 (D/D+G) parameter. Data shown for core (filled circles) and drill cuttings (open circles) and for predictions made by variable and constant heat flow models (solid and dashed lines) g) Vitrinite reflectance data and models.

Figure 4. a) Fraction of TOC converted to recalcitrant kerogen. Data is shown for Units II – V, and the y-axis is based on the fractional change in the abundance of sterenes (here scaled from 0-1). Line is polynomial equation described by Equation 6. b) Fraction of inorganic carbon in carbonate as fraction of total carbon. Data is shown for mudstones and siltstones in Units III and IV, black line represents Equation 7. c) DOC vs. TOC relationship used to empirically obtain the fraction of carbon that is bioavailable for the deep biosphere in thermally immature sediments. Black line represents Equation 8.

Figure 5. Downhole plots of carbon-loss functions and the data from which the functions were derived. a) Sterene concentrations (filled circles) and average values used for Figure 4a (open circles). b) Total inorganic carbon (filled circles) and average values (open circles). c) Fractional loss of organic carbon bound into recalcitrant kerogen -  $C_{lker}$  and d) Fractional loss of organic carbon sequestered in authigenic carbonate -  $C_{lcarb}$ . Figures c and d computed using the variable heat flow basin model in Figure 3 and Figure 4.

Figure 6. Comparison of the behaviour of material balance and population deep biosphere models. a) & b) Are hypothetical models with growth ratios of 1.1 (a) and 0.9 (b) in which the availability of organic carbon declines as shown in c).

Figure 7. Downhole predictions of cell abundances made by different combinations of Basin Models and Deep Biosphere Models. a) & b) Lithology and TOC content. c) Predictions of present-day cell abundances using a Basin Model with variable heat flow (VHF) and a Material Balance Model. d) Predictions of present-day cell abundances using a Basin Model with variable heat flow (VHF) and a Population Model. e) Predictions of present-day cell abundances using a Basin Model with constant heat flow (CHF) and a Material Balance Model. For c), d) and e) maximum and minimum cell concentrations are shown as predictions for the lowest and highest TOC lithologies within a formation.

Figure 8. Burial curves illustrating the development of distinctive biomes over deep geological time within Units I-V. Predictions of cell abundances are shown for the maximum- and minimum-TOC content of each unit for; a-b) the variable heat flow basin and material balance models, c-d) the constant heat flow basin and material balance models and e-f) variable heat flow basin and population models.

Figure 9. Plots illustrating the differing effects of cell-damage and starvation on cell abundances. The effects of differing geological history on cell abundances in Unit III. a) Cell abundances calculated by the constant heat flow and variable heat flow models. b) Rates of cell damage and temperature calculated for the constant heat flow and variable heat flow models, showing divergence at 5 Ma. d) Cell abundances calculated for the high and low TOC lithologies within Unit II. e) Changes in porosity and the fraction of organic carbon bound in carbon over geological time ( $C_{ker}$ ). f) Plot of the organic carbon available to support the deep biosphere ( $O_{Cr}$ ) in the high and low TOC lithologies within Unit II and the organic carbon needed for cell repair. Significant decreases in cell abundance occur when the carbon needed to support the biosphere is no longer available.

Figure 10. Plots of a) cell abundance and b) carrying capacity within the high TOC and low TOC lithologies in Unit IV (upper estuarine facies) calculated using the population model and variable heat flow basin model.

Figure 11. Downhole plot of a) S3 – yield of combustible/oxidised carbon per gram of rock, b) cell abundance predicted using a variable heat flow Basin Model and Population Biosphere Model with varied stimulation of the coalbed horizon biome by carbon dioxide generated from net thickness of coal seams 2, 20 and 40 m thick.

Figure 12. a) Burial curve showing the stimulation of a low TOC sandstone within the coalbed horizon where an overlying sandstone is stimulated by CO<sub>2</sub>. b) Varying stimulation as a function of the net thicknesses of coalbed, and the generation of carbon dioxide from a 20 meters net thickness of coalbed during the late Miocene to present day. c) The organic carbon available to the biosphere with and without the addition of CO<sub>2</sub>.

## References

- Athy, L. F., (1930). Density, porosity, and compaction of sedimentary rocks, *AAPG Bull.*, *14*, 1–22.
- Amend, J.P., Shock, E.L. (2001). Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS Microbiology Reviews*, *25*, 175-243. <http://dx.doi.org/10.1111/j.1574-6976.2001.tb00576.x>
- Bernard, F.P., Connan, J., (1992). Indigenous microorganisms in connate waters of many oilfields: a new tool in exploration and production techniques. *SPE 24811. In: 67th Annual Technical Conference and Exhibition of the Society of Petroleum Engineers, Washington, DC, October 1992*, 467-476.
- Bennett, B., Aitken, C.M., Jones, D.M., Farrimond, P., Larter, S.R. (2007). The occurrence and significance of 25-norhopanoic acids in petroleum reservoirs, *Organic Geochemistry*, *38* (11), 1977-1985. <http://dx.doi.org/10.1016/j.orggeochem.2007.06.011>
- Berner, R. A. (1968)- Calcium carbonate concentrations formed by the decomposition of organic matter. *Science* *159*, 195–197. <http://dx.doi.org/10.1126/science.159.3811.195>
- Beulig F., Røy, H., Glombitza C., and Jørgensen B. B. (2018). Control on rate and pathway of anaerobic organic carbon degradation in the seabed, *PNAS*, *115*, 367-372
- Breuker A., Köweker, G., Blazejak A., Schippers A. (2011) The deep biosphere in terrestrial sediments in the Chesapeake Bay area, Virginia, USA . *Front Microbiol* *2*: Article 156. <http://dx.doi.org/10.3389/fmicb.2011.00156>
- Burnham, A. and Sweeny JJ. (1989). A Simple kinetic model of petroleum formation and cracking. *Geochim. Cosmochim. Acta.*, *43*, 649-2657.
- Burnham A.K. and Braun R.L. (1990). Development of a detailed model of petroleum formation, destruction, and expulsion from lacustrine and marine source rocks. *Organic Geochemistry*, *16*, 27-39. [http://dx.doi.org/10.1016/0146-6380\(90\)90023-S](http://dx.doi.org/10.1016/0146-6380(90)90023-S)



Braun R. L., Burnham A. K. (1987). Analysis of chemical reaction kinetics using a distribution of activation energies and simpler models, *Energy & Fuels*, 1, 153-161.

<http://dx.doi.org/10.1021/ef00002a003>

Braun R.L. and Burnham A.K. (1988). R.L. Braun, A.K. Burnham KINETICS: a computer program to analyze chemical reaction data, Lawrence Livermore National Laboratory Report UCID-21588.

Braun, S., Morono, Y., Littmann S., Kuypers, M., Aslan H., Dong M., Jørgensen, B.B., and Lomstein B.A., (2016). Size and Carbon Content of Sub-seafloor Microbial Cells at Landsort Deep, Baltic Sea Journal: *Frontiers in Microbiology*, Volume 7:1375.

<http://dx.doi.org/10.3389/fmicb.2016.01375>

Colwell, F. S., Boyd, S., Delwiche, M. E., Reed, D. W., Phelps, T. J., & Newby, D. T. (2008). Estimate of biogenic methane production rates in deep marine sediments at Hydrate Ridge, Cascadia Margin. *Applied and Environmental Microbiology*, 74, 3444–3452.

<http://dx.doi.org/10.1128/AEM.02114-07>

Connan, J. (1984) Biodegradation Of Crude Oils In Reservoirs. *Advances in Petroleum Geochemistry*, 1, 299-335. <http://dx.doi.org/10.1016/B978-0-12-032001-1.50011-0>

Crill M. and Martens C.S. 1984. Methane production from bicarbonate and acetate in an anoxic marine sediment. *Gochimica. et Cosmochimica. Acta*, 50, 2089-2097.

Daniels L, Sparling R, and Sprott G.D. 1984. The Bioenergetics of Methanogenesis. *Biochimica et Biophysica Acta*, 768, 113-163. [http://dx.doi.org/10.1016/0304-4173\(84\)90002-8](http://dx.doi.org/10.1016/0304-4173(84)90002-8)

Egeberg, P.K., and Abdullah, M.I., (1990). The diagenetic factors controlling the dissolved organic carbon (DOC) in pore water from deep sea sediments (ODP Leg 113, Weddell Sea). In: Barker, P.F., Kennett, J.P., et al., Proc. ODP, Sci. Results, 113: College Station, TX (Ocean Drilling Program), 169–177.

Emerson, S., Jahnke, R., Bender, M., Froelich, P., Klinkhammer, G., Bowser, C., Setlock, G. (1980). Early diagenesis in sediments from the eastern equatorial Pacific, I. Pore water nutrient and carbonate results. *Earth and Planetary Science Letters*, 49, 57–80.

[http://dx.doi.org/10.1016/0012-821X\(80\)90150-8](http://dx.doi.org/10.1016/0012-821X(80)90150-8)

Farrimond, P., Taylor, A., Telnæs, N. (1998). Biomarker maturity parameters: The role of generation and thermal degradation. *Organic Geochemistry*, 29, 1181-1197.

[http://dx.doi.org/10.1016/S0146-6380\(98\)00079-5](http://dx.doi.org/10.1016/S0146-6380(98)00079-5)

Glombitza, C., Mangelsdorf, K., Horsfield, B. (2016). Differences in bitumen and kerogen-bound fatty acid fractions during diagenesis and early catagenesis in a maturity series of New Zealand coals. *International Journal of Coal Geology*, 153, 28–36.

<http://dx.doi.org/10.1016/j.coal.2015.11.009>

Gross, D., Bechtel, A., and Harrington, G.J. (2015). Variability in coal facies as reflected by organic petrological and geochemical data in Cenozoic coal beds offshore Shimokita (Japan)—IODP Exp. 337. *International Journal of Coal Geology*, 152, 3–79.

<http://dx.doi.org/10.1016/j.coal.2015.10.007>

Hay, W. W., Sloan, J. L. I. and Wold C. N. (1998), The mass/age distribution of sediments on the ocean floor and the global rate of loss of sediment, *J. Geophys. Res.*, 93(B12), 14,933–14,940.

Head, I.M., Jones, D.M., Larter, S.R. (2003). Biological activity in the deep subsurface and the origin of heavy oil. *Nature*, 426, 344-352. <http://dx.doi.org/10.1038/nature02134>

Hantschel, T, and Kauerauf, A. I. (2009). Fundamentals of Basin and Petroleum Systems Modeling. Springer Berlin. Pages 103-150.

Hinrichs, K.-U., Inagaki, F. Downsizing the deep biosphere (2012). *Science*, 338 (6104), 204-205. <http://dx.doi.org/10.1126/science.1229296>

Honda, S. (1985). Thermal structure beneath Tohoku, northeast Japan. *Tectonophysics*, 112 (1-4), 69-102. [http://dx.doi.org/10.1016/0040-1951\(85\)90173-8](http://dx.doi.org/10.1016/0040-1951(85)90173-8)

Hoshino, T., Morono, Y., Terada, T., Imachi, H., Ferdelman, T. G., & Inagaki, F. (2011). Comparative study of subseafloor microbial community structures in deeply buried coral fossils and sediment matrices from the Challenger Mound in the Porcupine Seabight. *Frontiers in Microbiology*, 2, 231.  
<http://dx.doi.org/10.3389/fmicb.2011.00231>

Hoshino, T., and Inagaki, F. (2019) Abundance and distribution of Archaea in the subseafloor sedimentary biosphere. *ISME J.*, 13, 227-231. <http://dx.doi.org/10.1038/s41396-018-0253-3>

Horsfield, B., Schenk, H.J., Zink, K., Ondrak, R., Dieckmann, V., Kallmeyer, J., Mangelsdorf, K., di Primio, R., Wilkes, H., Parkes, R.J., Fry, J., Cragg, B. (2006). Living microbial ecosystems within the active zone of catagenesis: Implications for feeding the deep biosphere. *Earth and Planetary Science Letters*, 246, 55-69.  
<http://dx.doi.org/10.1016/j.epsl.2006.03.040>

Horstad, I., Larter, S.R. (1997). Petroleum migration, alteration, and remigration within Troll field, Norwegian North Sea. *AAPG Bulletin*, 81, 222-248.

House, C. H., Cragg, B. A., Teske, A., & the Leg 201 Scientific Party. (2003). Drilling contamination tests during ODP leg 201 using chemical and particulate tracers. In S. L. D'Hondt, B. B. Jørgensen, D. J. Miller, et al. (Eds.), *Proc ODP int repts (201) ( 1–19)*.

Inagaki, F., Suzuki, M., Takai, K., Oida, H., Sakamoto, T., Aoki, K., et al. (2003). Microbial communities associated with geological horizons in coastal subseafloor sediments from the Sea of Okhotsk. *Applied and Environmental Microbiology*, 69, 7224–7235.  
<http://dx.doi.org/10.1128/AEM.69.12.7224-7235.2003>

Inagaki, F., Nunoura, T., Nakagawa, S., Teske, A., Lever, M., Lauer, A., et al. (2006). Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments on the Pacific Ocean Margin. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 2815–2820.

Inagaki, F., Hinrichs, K.-U., Kubo, Y., and the Expedition 337 Scientists, (2013). *Proc. IODP, 337*: Tokyo (Integrated Ocean Drilling Program Management International, Inc.).

Inagaki, F., Hinrichs, K.-U., Kubo, Y., Bowles, M.W., Heuer, V.B., Hong, W.-L., Hoshino, T., Ijiri, A., Imachi, H., Ito, M., Kaneko, M., Lever, M.A., Lin, Y.-S., Methé, B.A., Morita, S., Morono, Y., Tanikawa, W., Bihan, M., Bowden, S.A., Elvert, M., Glombitza, C., Gross, D., Harrington, G.J., Hori, T., Li, K., Limmer, D., Liu, C.-H., Murayama, M., Ohkouchi, N., Ono, S., Park, Y.-S., Phillips, S.C., Prieto-Mollar, X., Purkey, M., Riedinger, N., Sanada, Y., Sauvage, J., Snyder, G., Susilawati, R., Takano, Y., Tasumi, E., Terada, T., Tomaru, H., Trembath-Reichert, E., Wang, D.T., Yamada, Y. Exploring deep microbial life in coal-bearing sediment down to ~2.5 km below the ocean floor (2015). *Science*, 349, 420-424.  
<http://dx.doi.org/10.1126/science.aaa6882>

Iversen, N., Jorgensen, B.B. Anaerobic methane oxidation rates at the sulfate - methane transition in marine sediments from Kattegat and Skagerrak (Denmark) (1985). *Limnology and Oceanography*, 30, 944-955. <http://dx.doi.org/10.4319/lo.1985.30.5.0944>

Jolivet, L., K. Tamaki, and M. Fournier (1994), Japan Sea, opening history and mechanism: A synthesis, *J. Geophys. Res.*, 99(B11), 22237–22259, <http://dx.doi.org/10.1029/93JB03463>

Jones, D.M., Head, I.M., Gray, N.D., Adams, J.J., Rowan, A.K., Aitken, C.M., Bennett, B., Huang, H., Brown, A., Bowler, B.F.J., Oldenburg, T., Erdmann, M., Larter, S.R. Crude-oil biodegradation via methanogenesis in subsurface petroleum reservoirs (2008). *Nature*, 451, 176-180. <http://dx.doi.org/10.1038/nature06484>

Kallmeyer, J., Smith, D. C., Spivack, A. J., & D'Hondt, S. (2008). New cell extraction procedure applied to deep subsurface sediments. *Limnology and Oceanography:Methods*, 6, 236–245.

Kallmeyer, J., Pockalny, R., Adhikari, R.R., Smith, D.C., D'Hondt, S. Global distribution of microbial abundance and biomass in subseafloor sediment (2012) *Proceedings of the National Academy of Sciences of the United States of America*, 109, 16213-16216.

Knoll, A.H., Bambach, R.K., Payne, J.L., Pruss, S., Fischer, W.W. Paleophysiology and end-Permian mass extinction (2007) *Earth and Planetary Science Letters*, 256, 295-313.

Kohnen, M.E.I., Sinninghe Damsté, J.S., Kock-van Dalen, A.c., Jan, W.D.L. (1991). Di- or polysulphide-bound biomarkers in sulphur-rich geomacromolecules as revealed by selective chemolysis. *Geochimica et Cosmochimica Acta*, 55, 1375-1394.

[http://dx.doi.org/10.1016/0016-7037\(91\)90315-V](http://dx.doi.org/10.1016/0016-7037(91)90315-V)

Lamorde, U.A., Parnell, J., Bowden, S.A. (2015). Constraining the genetic relationships of 25-norhopanes, hopanoic and 25-norhopanoic acids in onshore Niger Delta oils using a temperature-dependent material balance. *Organic Geochemistry*, 79, 31-43.

<http://dx.doi.org/10.1016/j.orggeochem.2014.12.004>

Larter, S., Wilhelms, A., Head, I., Koopmans, M., Aplin, A., Di Primio, R., Zwach, C., Erdmann, M., Telnaes, N. (2003). The controls on the composition of biodegraded oils in the deep subsurface - Part 1: Biodegradation rates in petroleum reservoirs. *Organic Geochemistry*, 34, 601-613. [http://dx.doi.org/10.1016/S0146-6380\(02\)00240-1](http://dx.doi.org/10.1016/S0146-6380(02)00240-1)

Langerhuus, A. T., Røy, H., Lever, M. A., Morono, Y., Inagaki, F., Jørgensen, B. B., et al. (2012). Endospore abundance and D: L-amino acid modeling of bacterial turnover in Holocene marine sediment (Aarhus Bay). *Geochimica et Cosmochimica Acta*, 99, 87–99.

<http://dx.doi.org/10.1016/j.gca.2012.09.023>

Lever, M. A. (2012). Acetogenesis in the energy-starved deep biosphere—a paradox?

*Frontiers in Microbiology*, 2, 284. <http://dx.doi.org/10.3389/fmicb.2011.00284>

Lever, M. A., Rogers, K. L., Lloyd, K. G., Overmann, J., Schink, B., Thauer, R. K., Hoehler, T. M., and Jørgensen, B. B. (2015). Life under extreme energy limitation: a synthesis of laboratory- and field-based investigations. *FEMS Microbiology Review*, 39, 688-728.

<http://dx.doi.org/10.1093/femsre/fuv020>

Lipp, J. S., Morono, Y., Inagaki, F., Hinrichs, K.-U. (2008). Significant contribution of Archaea to extant biomass in marine subsurface sediments. *Nature* 454, 991–994.

<http://dx.doi.org/10.1038/nature07174>

Liu, C.-H., Huang, X., Xie, T.-N., Duan, N., Xue, Y.-R., Zhao, T.-X., Lever, M. A., Hinrichs, K.-U., and Inagaki, F. (2017). Exploration of cultivable fungal communities in deep coal-bearing sediments from ~1.3 to 2.5 km below the ocean floor. *Environmental Microbiology*, *19*, 803-818.

Lomstein, B.A., Langerhuus, A.T., D'Hondt, S., Jørgensen, B.B., Spivack, A.J. (2012) Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. *Nature*, *484*, 101-104.

Magnabosco, C., Lin, L.-H., Dong, H. , Bomberg, M. , Ghiorse, W., Stan-Lotter, H., Pedersen, K., Kieft, T. L., van Heerden E., Onstott, T. C. (2018). The biomass and biodiversity of the continental subsurface. *Nature Geoscience* *11*, 707–717.  
<http://dx.doi.org/10.1038/s41561-018-0221-6>

McKenzie, D. Some remarks on the development of sedimentary basins (1978). *Earth and Planetary Science Letters*, *40*, 25-32. [http://dx.doi.org/10.1016/0012-821X\(78\)90071-7](http://dx.doi.org/10.1016/0012-821X(78)90071-7)

McMahon, S. and Parnell J. (2014) Weighing the deep continental biosphere. *FEMS Microbiology Ecology*, *87*, 113–120. <http://dx.doi.org/10.1111/1574-6941.12196>

Meister, P. (2015). For the deep biosphere, the present is not always the key to the past: What we can learn from the geological record. *Terra Nova*, *27*, 400-408.  
<http://dx.doi.org/10.1111/ter.12174>

Morita, R. Y., & ZoBell, C. L. (1955). Occurrence of bacteria in pelagic sediments collected during the Mid-Pacific expedition. *Deep-Sea Research*, *3*, 66–73.  
[http://dx.doi.org/10.1016/0146-6313\(55\)90036-8](http://dx.doi.org/10.1016/0146-6313(55)90036-8)

Morono, Y., Terada, T., Masui, N., and Inagaki, F., (2009). Discriminative detection and enumeration of microbial life in marine subsurface sediments. *ISME J.*, *3*, 503–511.  
<http://dx.doi.org/10.1038/ismej.2009.1>

Middelburg J. J. (1989) A simple rate model for organic matter decomposition in marine sediments. *Geochim. Cosmochim. Acta* 53, 1577–1581.

[http://dx.doi.org/10.1016/0016-7037\(89\)90239-1](http://dx.doi.org/10.1016/0016-7037(89)90239-1)

Muirhead, D.K., Parnell, J., Taylor, C., Bowden, S.A. (2012). A kinetic model for the thermal evolution of sedimentary and meteoritic organic carbon using Raman spectroscopy. *Journal of Analytical and Applied Pyrolysis*, 96, 153-161.

<http://dx.doi.org/10.1016/j.jaap.2012.03.017>

Nunoura, T., Soffientino, B., Blazejak, A., Kakuta, J., Oida, H., Schippers, A., & Takai, K. (2009). Subseafloor microbial communities associated with rapid turbidite deposition in the Gulf of Mexico continental slope (IODP Expedition 308). *FEMS Microbiology Ecology*, 69, 410–424. <http://dx.doi.org/10.1111/j.1574-6941.2009.00718.x>

Omar G Onstott T Hoek J (2003). The origin of deep subsurface microbial communities in the Witwatersrand Basin, South Africa as deduced from apatite fission track analyses.

*Geofluids* 3: 69–80. <http://dx.doi.org/10.1046/j.1468-8123.2003.00050.x>

Parkes, R.J., Cragg, B.A., Bale, S.J., Getliff, J.M., Goodman, K., Rochelle, P.A., Fry, J.C., Weightman, A.J., Harvey, S.M. (1994). Deep bacterial biosphere in Pacific Ocean sediments. *Nature*, 371, 410-413. <http://dx.doi.org/10.1038/371410a0>

Parkes, R.J., Cragg, B.A., Wellsbury, P. (2000). Recent studies on bacterial populations and processes in subseafloor sediments: A review. *Hydrogeology Journal*, 8, 11-28.

<http://dx.doi.org/10.1007/PL00010971>

Peters, K. E., Walters, C. C. and J. M. Moldowan (2005). The Biomarker Guide, Vol 2. Cambridge University Press.

Peters, K.E., Nelson, P.H. (2012). Analysing the Thermal History of Sedimentary Basins: Methods and Case Studies, *SEPM Special Publication* 103, 5-15.

Phillips, S. C., Johnson, J. E., Clyde, W. C., Setera, J. B., Maxbauer, D. P., Severmann, S., and Riedinger, N. (2017) .Rock magnetic and geochemical evidence for authigenic magnetite



formation via iron reduction in coal - bearing sediments offshore Shimokita Peninsula, Japan (IODP Site C0020). *Geochem. Geophys. Geosyst.*, 18, 2076–2098.

<http://dx.doi.org/10.1002/2017GC006943>

Phillips, S. C., Hong, W.-L., Johnson, J. E., Fahnstock, M. F., & Bryce, J. G. (2018).

Authigenic carbonate formation influenced by freshwater inputs and methanogenesis in coal-bearing strata offshore Shimokita, Japan (IODP Site C0020). *Marine and Petroleum Geology*, 96, 288–303. <http://dx.doi.org/10.1016/j.marpetgeo.2018.06.007>

Price, P. B, Sowers T. (2004). Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. *Proceedings of the National Academy of Sciences* 101, 4631-4636.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Rabinowitz, H. S., P. J. Polissar, and H. M. Savage (2017). Reaction kinetics of alkenone and *n*-alkane thermal alteration at seismic timescales, *Geochem. Geophys. Geosyst.*, 18, 204-219.

<http://dx.doi.org/10.1002/2016GC006553>

Romanova, N.D. & Sazhin, A.F. (2010). Relationships between the cell volume and the carbon content of bacteria. *Oceanology* 50, 522.

<http://dx.doi.org/10.1134/S0001437010040089>

Roussel, E. G., Cragg, B. A., Webster, G., Sass, H., Tang, X., Williams, A. S., ... Parkes, R. J. (2015). Complex coupled metabolic and prokaryotic community responses to increasing temperatures in anaerobic marine sediments: critical temperatures and substrate changes. *FEMS Microbiology Ecology*, 91, fiv084. <http://doi.org/10.1093/femsec/fiv084>

Renshaw, Eric (1991). Modeling Biological Populations in Space and Time. Cambridge University Press. 6–9.

Sweeney, J.J., Burnham, A.K. (1990) Evaluation of a simple model of vitrinite reflectance based on chemical kinetics. *American Association of Petroleum Geologists Bulletin*, 74, 1559-1570.

Sajgó C., Lefler J. (1986). A reaction kinetic approach to the temperature-time history of sedimentary basins. In: Buntebarth G., Stegena L. (eds) *Paleogeothermics. Lecture Notes in Earth Sciences*, vol 5. Springer, Berlin, Heidelberg.

Tanikawa, W., Osamu Tadai, Sumito Morita, Weiren Lin, Yasuhiro Yamada, Yoshinori Sanada, Kyaw Moe, Yu'suke Kubo, Fumio Inagaki, (2016). Thermal properties and thermal structure in the deep-water coalbed basin off the Shimokita Peninsula, Japan. *Marine and Petroleum Geology*, 73, Pages 445-461. <http://dx.doi.org/10.1016/j.marpetgeo.2016.03.006>

Tanikawa, W., Tadai, O., Morono, Y., Hinrichs, K.-U., and Inagaki, F. (2018). Geophysical constraints on microbial biomass in subseafloor sediments down to ~2.5 km-deep coal seams off Shimokita Peninsula, Japan. PEPS, (under Revision).

Tegelaar E. W., and Noble R.K. (1994). Kinetics of hydrocarbon generation as a function of the molecular structure of kerogen as revealed by pyrolysis-gas chromatography, *Organic Geochemistry*, 22, 543-574. [http://dx.doi.org/10.1016/0146-6380\(94\)90125-2](http://dx.doi.org/10.1016/0146-6380(94)90125-2)

Tissot, B.P., Welte, D.H. (1984). *Petroleum Formation and Occurrence* Authors:, Springer Verlag.

Trembath-Reichert, E., Morono, Y., Ijiri, A., Hoshino, T., Dawson, K. S., Inagaki, F., & Orphan, V. J. (2017). Methyl-compound use and slow growth characterize microbial life in 2-km-deep subseafloor coal and shale beds. *Proceedings of the National Academy of Sciences*, 114, E9206–E9215.

van Krevelen, D.W. (1950). Graphical-statistical method for the study of structure and reaction processes of coal, *Fuel*, 29, 269-84.

Volkman, J.K., Alexander, R., Kagi, R.I., Woodhouse, G.W. (1983) Demethylated hopanes in crude oils and their applications in petroleum geochemistry. *Geochimica et Cosmochimica Acta*, 47, 785-794. [http://dx.doi.org/10.1016/0016-7037\(83\)90112-6](http://dx.doi.org/10.1016/0016-7037(83)90112-6)

Wangersky, P. (1978). Lotka-Volterra Population Models. *Annual Review of Ecology and Systematics*, 9, 189-218. <http://dx.doi.org/10.1146/annurev.es.09.110178.001201>

Waples, D. W. , (1980). Time and temperature in petroleum formation: application of Lopatin's method to petroleum exploration. *AAPG Bulletin*, 64, 916-926.

Waples, D.W., (1994). Modeling of sedimentary basins and petroleum systems .The petroleum system - from source to trap, 307-322.

Webster, G., John Parkes, R., Cragg, B. A., Newberry, C. J., Weightman, A. J., & Fry, J. C. (2006). Prokaryotic community composition and biogeochemical processes in deep seafloor sediments from the Peru Margin. *FEMS Microbiology Ecology*, 58, 65–85. <http://dx.doi.org/10.1111/j.1574-6941.2006.00147.x>

Weinbauer, M.G., Beckmann,C., Höfle M.G. (1998). Utility of Green Fluorescent Nucleic Acid Dyes and Aluminum Oxide Membrane Filters for Rapid Epifluorescence Enumeration of Soil and Sediment Bacteria. *Appl. Environ. Microbiol.* 64, 5000-5003.

Westrich, J. and Berner R. A. (1984). The role of sedimentary organic matter in bacterial sulfate reduction: The G-model tested. *Limnology and Oceanography*,29, 236 -249. <http://dx.doi.org/10.4319/lo.1984.29.2.0236>

Wilhelms, A., Larter, S.R., Head, I., Farrimond, P., Di-Primio, R., Zwach, C. (2001). Biodegradation of oil in uplifted basins prevented by deep-burial sterilization *Nature*, 411, 1034-1037. <http://dx.doi.org/10.1038/35082535>

Wilkinson, M. & Damper, M. D. (1990). The rate of growth of sandstone-hosted calcite concretions. *Geochim. Cosmochim. Acta* 54, 3391–3399. [http://dx.doi.org/10.1016/0016-7037\(90\)90293-T](http://dx.doi.org/10.1016/0016-7037(90)90293-T)

Wopenka B. and Pasteris, J.D. (1993). Structural characterization of kerogens to granulite-facies graphite: Applicability of Raman microprobe spectroscopy. *American Mineralogist* 78, 533-557.

Zeikus JG. (1977). The biology of methanogenic bacteria. *Bacterio Rev* 41, 514-41.

Zhou, Q., Xiao, X., Pan, L., Tian, H., (2014). The relationship between micro-Raman spectral parameters and reflectance of solid bitumen. *International Journal of Coal Geology*, 121, 19-25. <http://dx.doi.org/10.1016/j.coal.2013.10.013>

Table 1. Geochemical properties of modelled units

Unit	Top	Bottom	Lithologies	Age	Drill Cuttings Lithology			TOC		Rock Eval					Other		
					%			%		S2	S3	HI	OI	Tmax	TIC	C <sub>29</sub> sterene	C <sub>29</sub> sterane
					Mud	Silt	Sand	Low	high	mg HC/g	mgCO <sub>2</sub> /g	mg HC/g TOC	mg CO <sub>2</sub> /g TOC	°C	%	ug/gTOC	ug/gTOC
OB	1180	1816	Over burden (Marine Mud)	Quaternary	nd	nd	nd	0.6	0.6	-	-	-	-	-	-	-	-
Unit V	1816	2436	Marine Muds, partially indurated at base with sand units	Pliocene	53	37	10	0.08	3.2	2.5	1.2	292	113	405	0.09	0.003	0.0009
Unit IV	2436	2680	Estuarine Siltstones and Mudstones, coal fragments, coaly sandstones		33	53	14	0.1	0.9	0.7	0.7	80	167	413	0.12	0.003	0.0004
Unit III	2680	3006	Estuarine mudstones with sandstones at base. Plant macrophytes common		52	33	15	0.1	1.0	1.0	0.9	80	178	417	0.15	0.002	0.0004
Unit II	3006	3226	Sandstones and Coalbeds	Early to Mid Miocene	29	52	20	0.07	45	95	14	245	125	428	0.20	0.001	0.0003
Unit I	3226	3646	Mudstones, coalbeds at base, plant macrophytes common	Late Olig	55	29	17	0.1	1.2	0.5	0.7	52	156	426	0.27	0.0004	0.0003

Table 2. Physical properties of modelled units

Unit	Top	Bottom	Lithology	Petromod Lithology	Thermal conductivities	Formation Density	Porosity
	meters subsea				mW.m <sup>-2</sup>	g.cm <sup>-3</sup>	%.
OB	1180	1816	Over burden (marine mud)	Shale	1.11	1.81	54
Unit V	1816	2436	Marine muds, partially indurated at base with sand units	Siltstone	1.63	2.11	36
Unit IV	2436	2680	Estuarine siltstones and mudstones, coal fragments, coaly sandstones	Sandstone	2.12	2.28	28
Unit III	2680	3006	Estuarine mudstones with sandstones at base. Plant macrophytes common	Siltstone	1.74	2.30	26
Unit II	3006	3226	Sandstones and coalbeds	70 % Sandstone & 30 % Coal	1.96	2.05	11
Unit I	3226	3646	Mudstones, coalbeds at base, plant macrophytes common	Siltstone	1.78	2.37	20

Table 3 - Parameters

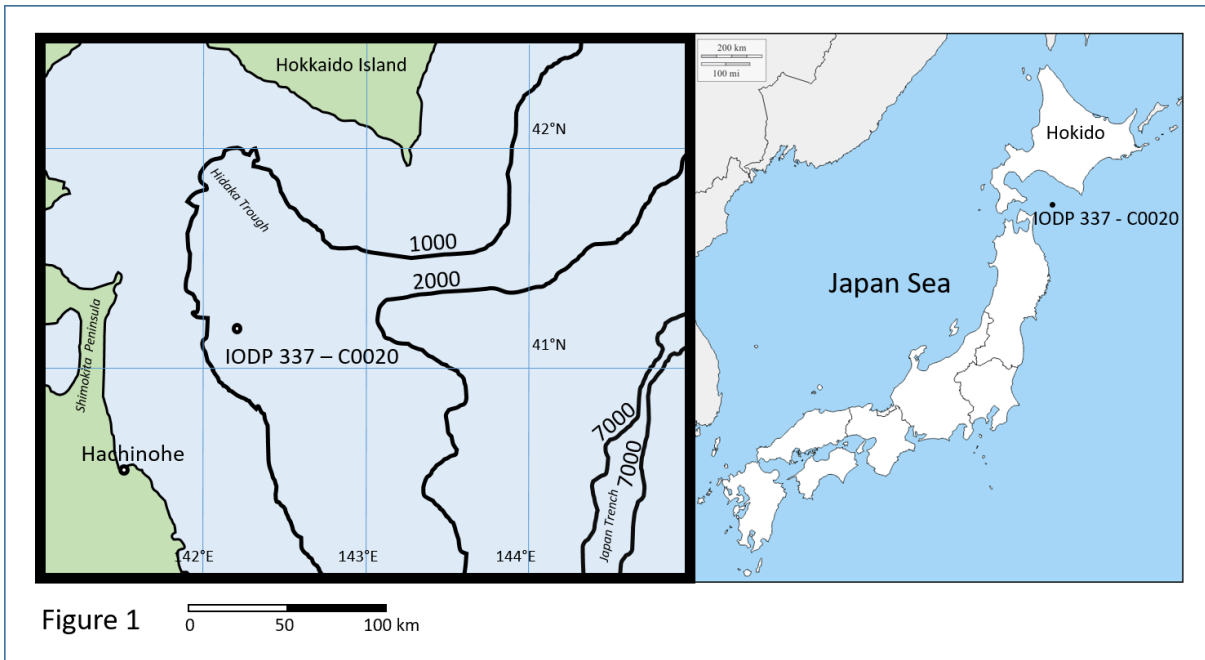
Parameter	Description	Units	Notes
$N_i, N_0$	Organic carbon in active deep biosphere at time steps 0 and i	grams carbon (gC. cm <sup>-3</sup> )	Equation 1 and Table 3 for initial conditions
$N_d$	Organic carbon lost from deep biosphere	grams carbon (gC. cm <sup>-3</sup> )	Equation 2
$N_g$	Organic carbon incorporated into deep biosphere	grams carbon (gC. cm <sup>-3</sup> )	Equation 3
$T$	Temperature	°C	calculated from basin model or measured
$t$	duration of time step	Seconds (s)	Set by user
$k_{dam}$	rate of cell damaging process	gC of a biomolecule per unit time (gC.s <sup>-1</sup> )	Price and Sowers 2004
$F_d$	weighting factor for cell damaging process	Dimensionless unit 0 -1	Set by user. (See Table 4)
$k_{rep}$	rate of cell repair process	gC of a biomolecule per unit time (gC.s <sup>-1</sup> )	Price and Sowers 2004
$F_{rep}$	weighting factor for cell repair process	Dimensionless unit 0 -1	Set by user (See Table 4)
$D_{oc}$	Bio-available organic carbon	Dimensionless unit 0 -1	Equation 8
$m_{DoC}$	Gradient describing relationship between DOC and TOC in low TOC sediments	1/TOC	By regression from data in Figure 4
$C_{DoC}$	Intercept describing relationship between DOC and TOC in low-TOC sediments	Dimensionless unit 0-1	By regression from data in Figure 4
$O_{cr}$	Organic carbon not kerogen-bound, not sequestered in carbonate, not in active deep biosphere	gC. cm <sup>-3</sup>	Equation 3
$TOC$	Total organic carbon	% Divided by 100 for computation	Values from basin Model. Present day - Table 1
$\phi$	Porosity	% Divided by 100 for computation	Values from basin Model. Present day - Table 1
$\rho$	Density	g.cm <sup>-3</sup>	Values from basin Model. Present day - Table 1
$C_{loss}$	Loss of carbon from bioavailable pool	Dimensionless unit from 0 -1	Equation 5
$C_{ker}$	Loss of carbon to kerogen	Dimensionless unit 0 -1	Equation 6
$V_{RCalc}$	Vitrinite Reflectance Calculated by Basin Model	% light reflected	Obtained from basin model. Present day - Table 1
$a_{ker}$	Exponent describing polynomial relationship between Vitrinite Reflectance Calculated by Basin Model and $C_{ker}$		By regression from data in Figure 4
$b_{ker}$	Exponent describing polynomial relationship between Vitrinite Reflectance Calculated by Basin Model and $C_{ker}$		By regression from data in Figure 4
$c_{ker}$	Intercept describing relationship between vitrinite reflectance and $C_{ker}$		By regression from data in Figure 4
$C_{carb}$	Loss of carbon to carbonate	Dimensionless unit 0 -1	Equation 7
$m_{lcarb}$	Gradient describing relationship between porosity and authigenic carbonate	1/log $\phi$	By regression from data in Figure 4
$C_{lcarb}$	Intercept describing relationship between porosity and authigenic carbonate	Dimensionless unit 0 -1	By regression from data in Figure 4
$a$	growth ratio	Dimensionless unit	Equation 10
$cc$	carrying capacity	gC. cm <sup>-3</sup>	Equation 11

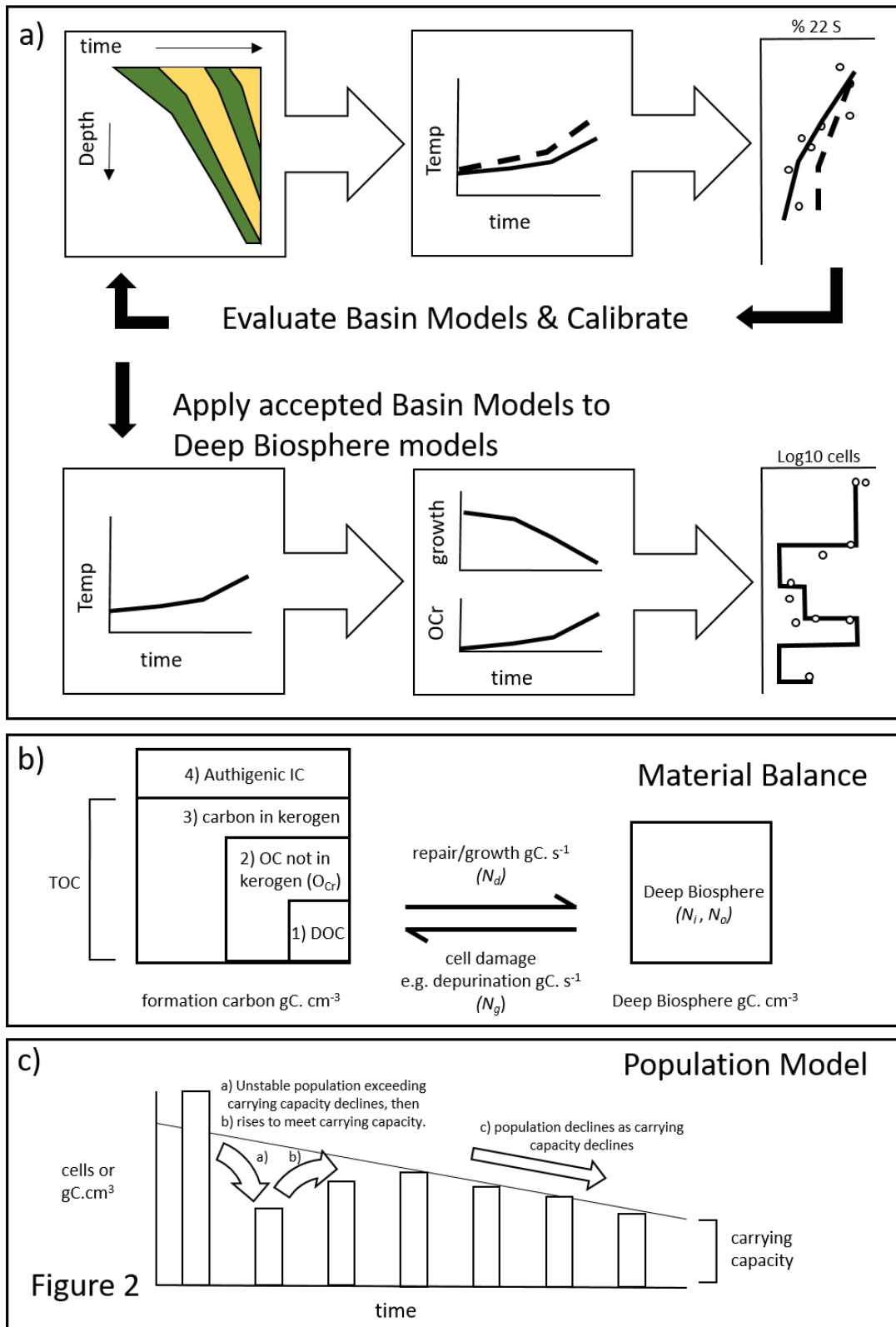


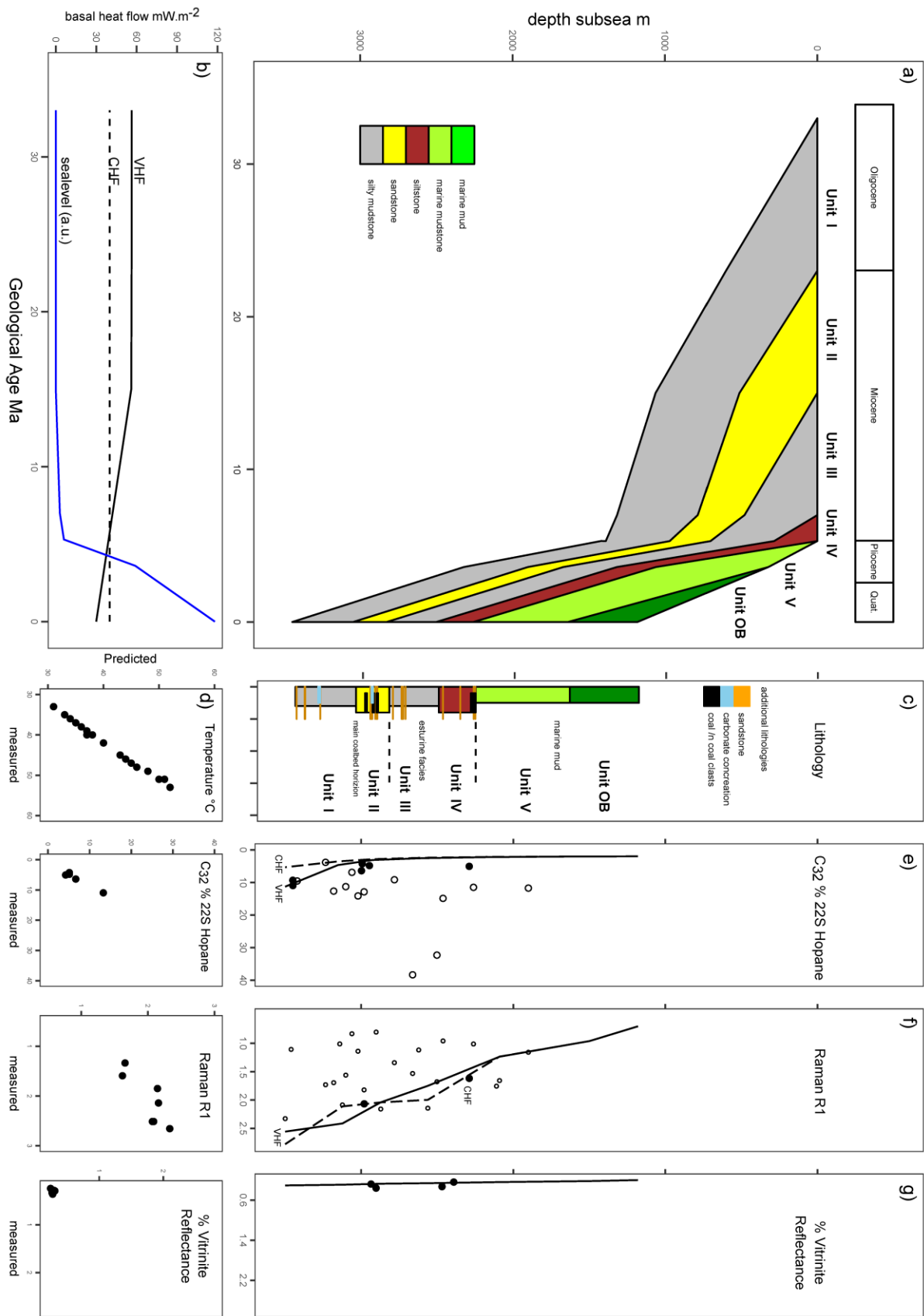
Table 4. Input and Initialisation Data for Deep Biosphere Models

		Unit	Equation	Reference (Data Source)
cellular carbon content	55 E-12	gC.cell <sup>-1</sup>	-	Romanova & Sazhin 2010
initial cell concentration	4.5 E 8	cells.cm <sup>-3</sup>	-	Inagaki et al 2015
Exponents obtained from regression of data				
cell damage to racemisation $m_{dam-rac}$	0.076	gC(gC) <sup>-1</sup> .h <sup>-1</sup> .(1000.K <sup>-1</sup> )	Eq. 2 & 11	Price and Sowers 2004
cell damage to racemisation $c_{dam-rac}$	28	gC(gC) <sup>-1</sup> .h <sup>-1</sup>	Eq. 2 & 11	
cell damage to racemisation ( $F_{dam-rac}$ )	0.25	a.u.	Eq. 2 & 11	
cell damage to depurination $m_{dam-dep}$	0.097	gC(gC) <sup>-1</sup> .h <sup>-1</sup> .(1000.K <sup>-1</sup> )	Eq. 2 & 11	
cell damage to depurination $c_{dam-dep}$	36.	gC(gC) <sup>-1</sup> .h <sup>-1</sup>	Eq. 2 & 11	
cell damage to depurination ( $F_{dam-dep}$ )	0.75	a.u.	Eq. 2 & 11	
cell maintenance $m_{rep}$	20	gC(gC) <sup>-1</sup> .h <sup>-1</sup> .(1000.K <sup>-1</sup> )	Eq. 3 & 11	
cell maintenance $c_{rep}$	120	gC(gC) <sup>-1</sup> .h <sup>-1</sup>	Eq. 3 & 11	
loss to carbonate $m_{carb}$	-0.053	% TIC/TC/φ	Eq. 7	Inagaki et al 2013
loss to carbonate $c_{carb}$	0.22	% TIC/TC	Eq. 7	
loss to kerogen $a_{ker}$	-6.8	% TOC/% VR	Eq. 6	Inagaki et al 2013
loss to kerogen $b_{ker}$	7.3	% TOC/% VR	Eq. 6	
loss to kerogen $c_{ker}$	-1.2	% TOC	Eq. 6	
bioavailable organic carbon ( as DOC) $D_{aC}$	0.0016	% DOC/% TOC	Eq. 8	Egeburg & Abdullah 1900

The parameters for cell-damaging reactions have additional subscript notation denoting the reaction concerned, e.g. for depurination this is  $_{dep}$  and for racemisation this is  $_{rac}$ . gC = grams Carbon; TIC = total inorganic carbon; TC = total carbon; TOC = total organic carbon; DOC = dissolved organic carbon; VR = vitrinite reflectance.







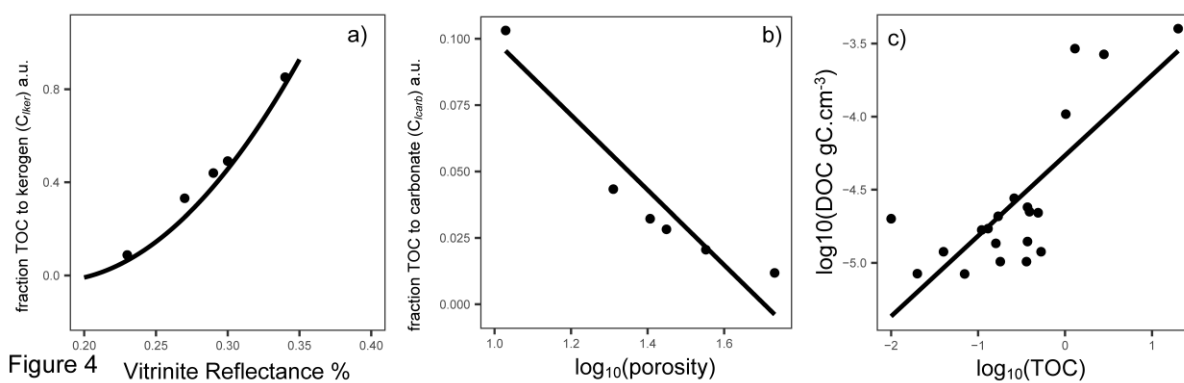
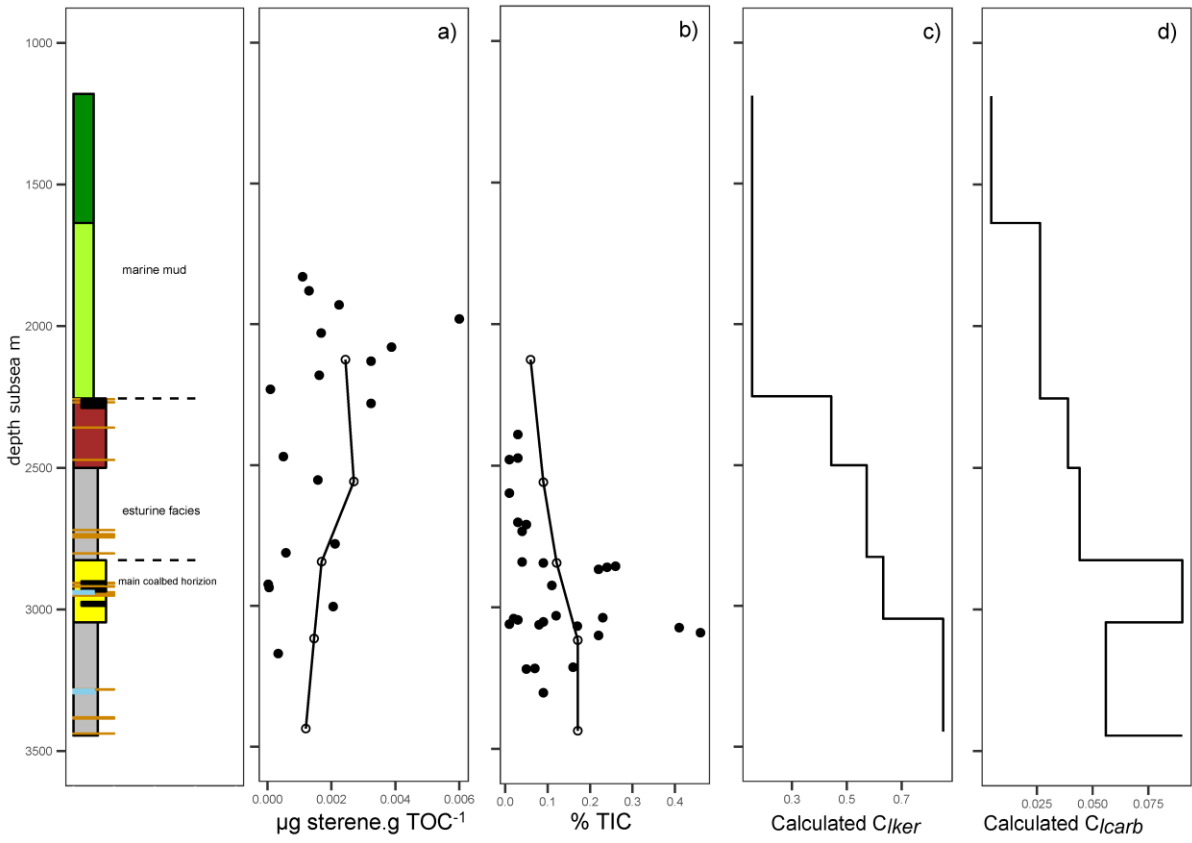


Figure 4 Vitrinite Reflectance %



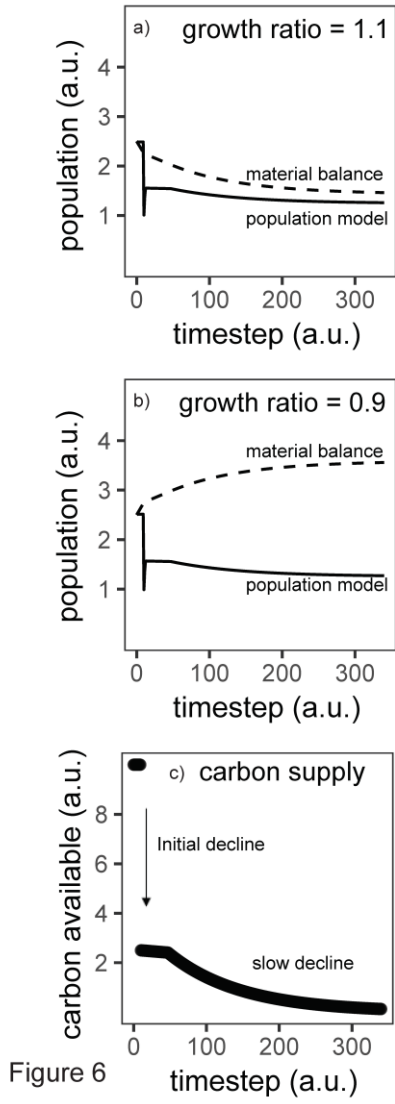
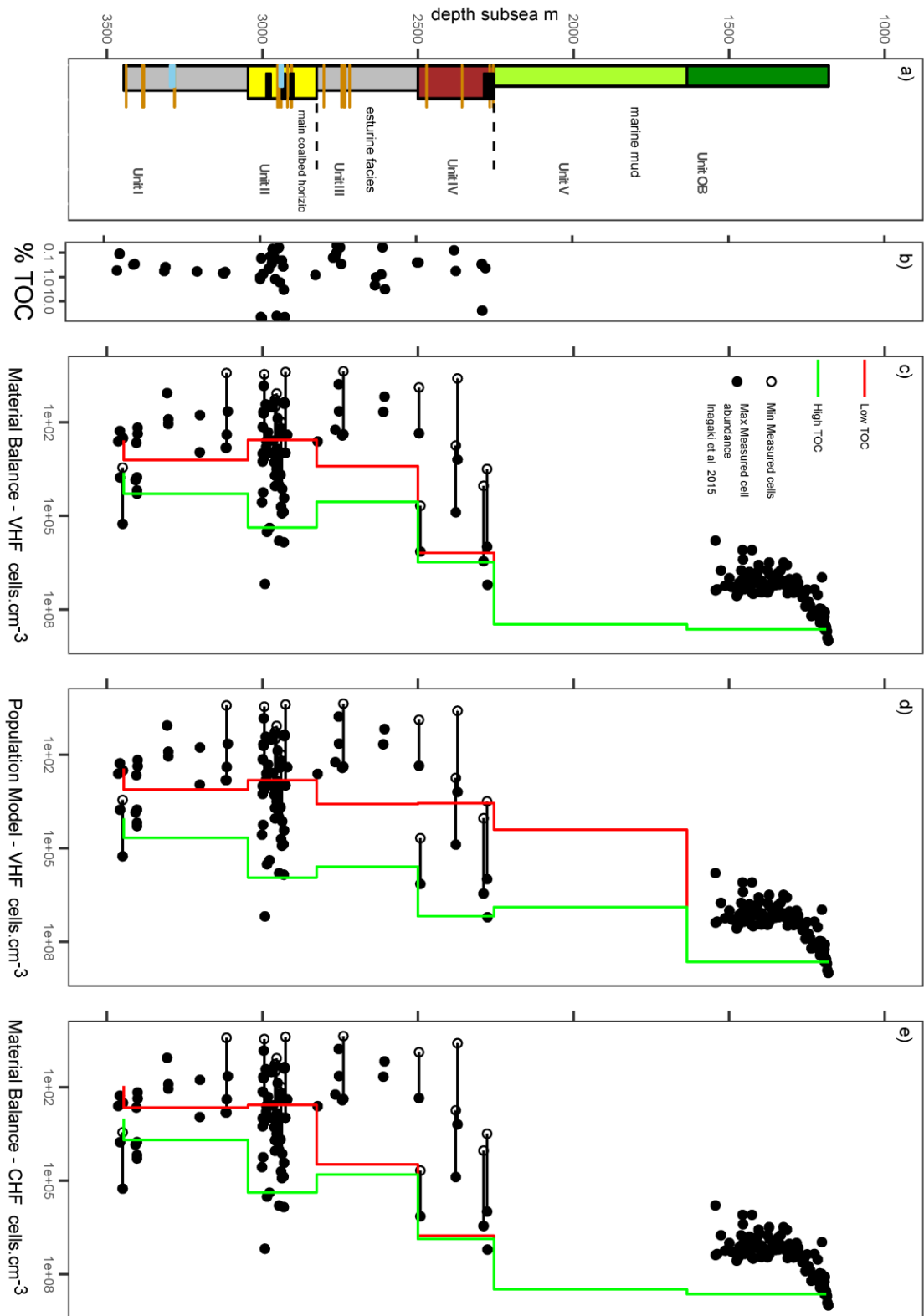
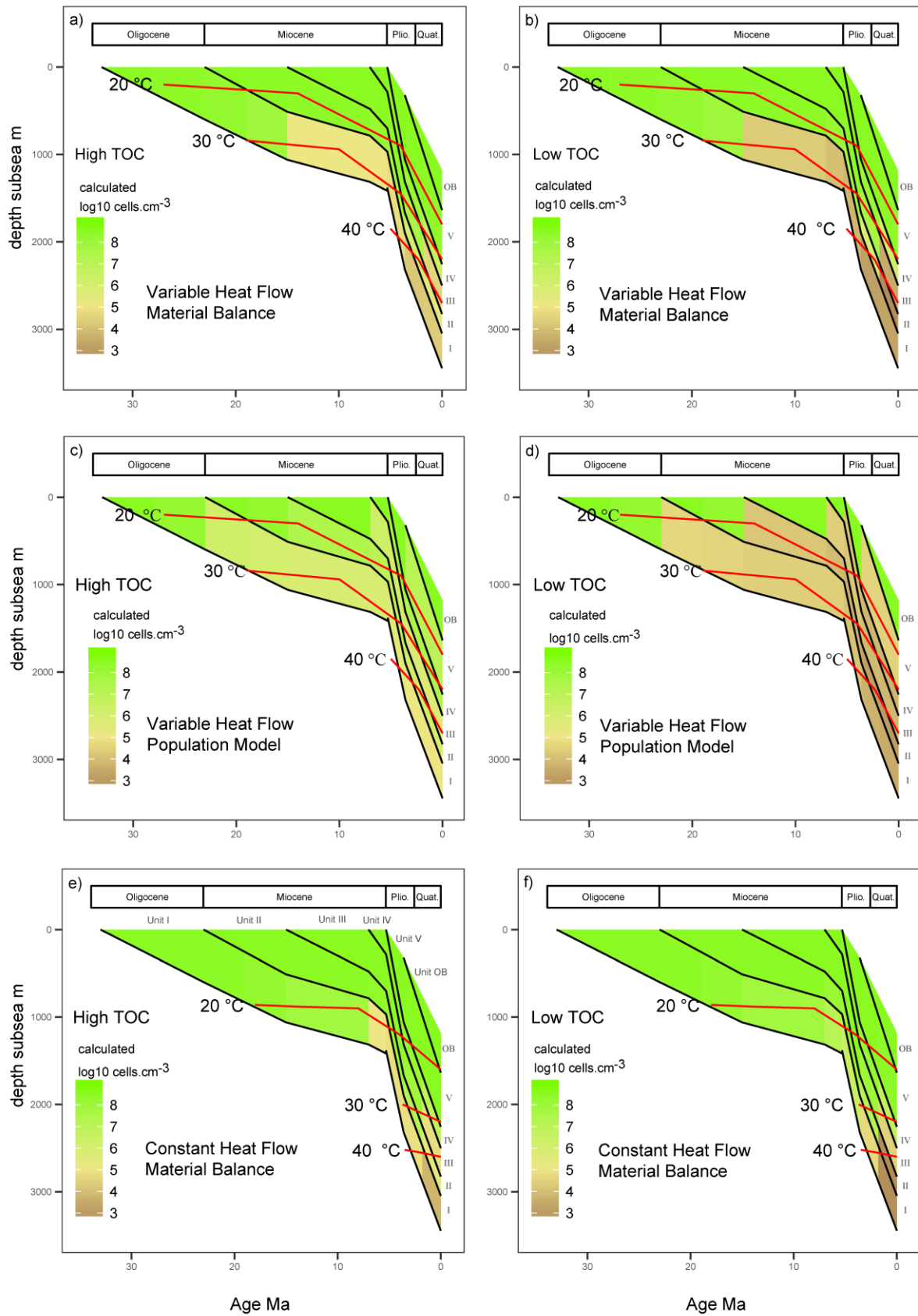


Figure 6







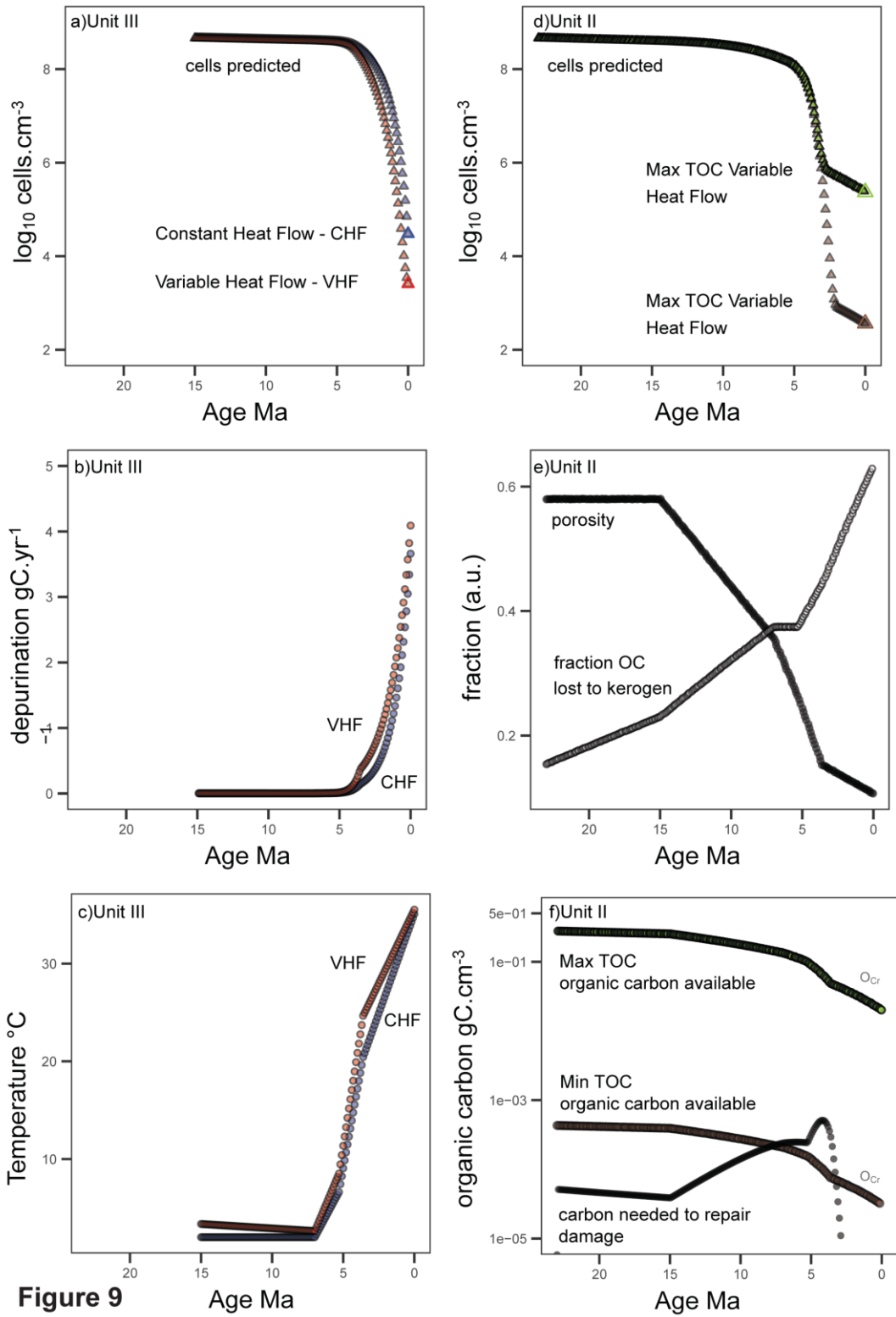


Figure 9

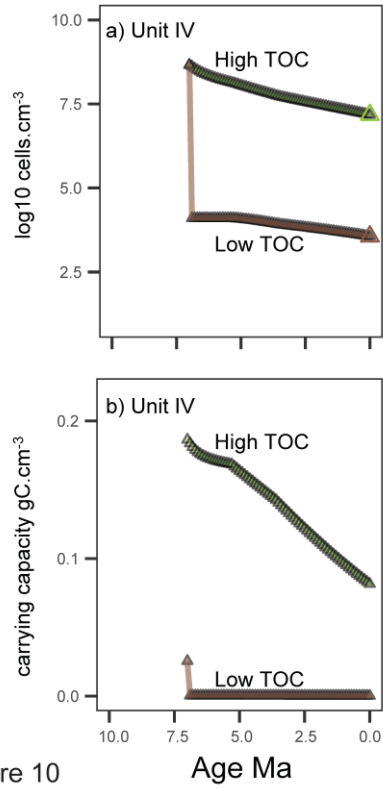


Figure 10

