Navigation tools for complex systems:
Seamlessness, rootedness and constraint resolution as aids to pattern-oriented modelling - *Insights from ecology.*

P. D. Chapman

1 Introduction

In his (2009) paper Chapman struggled with the evidence that his demonstrably imperfect model was explaining the experimental data with a very high regression coefficient ($r^2 > 0.99$), implying some sort of perfection not inherent in the derivation. This paper argues that this apparent dissonance is an example of a model that is in the (ecologically inspired notion of a) Medawar zone (Grimm, et al., 2005). Indeed the micro-environments of Chapman (2008) are clearly synonymous with Grimm's *patterns*.

This paper builds on the work of Grimm et al. and argues that the patterns of their pattern-oriented modelling can be formed using the tools of seamlessness (this being the 'voice' of the underlying causation) and constraint resolution. When patterns are formed in this manner the model structure is optimised and thereby located in the Medawar zone. Seamlessness plays a role in identifying modelling constraints that need to be resolved as bottom-up modelling is applied to ever widening contexts (discussed further in Chapman (2011)) leading to identification of patterns of similar behaviour (or composting rate in the composting case) that preserve the underlying causation.

When using seamlessness in a mathematical formulation the information inherent in *all* of the contributing parts is carried into the structure that is formed. This information carrying characteristic of mathematics becomes very useful for building models that are explaining complex systems.

2 Maintaining causation

In terms of causation in composting it could be argued that: microbes + electron acceptor + substrate will cause degradation of the substrate. These are *necessary* and *sufficient* conditions, although substrate in this context would need to include the essential nutrients. This is chemistry orchestrated by microbial activity; change the substrate or the electron acceptor and the rate of degradation will change. Similarly remove the microbes (or limit their numbers) and degradation will stop (or slow down).

This is the fundamental unit for modelling a microbial system such as composting, and exists at a small scale: a cell wall will degrade differently from the cell contents due to it being a different substrate; oxygen diffuses only mm into a composting particle. Beyond the reach of oxygen a different electron acceptor will be used.

To maintain causation, a model must accommodate these differences. It follows that for a model describing microbial degradation of a substrate, then to maintain causation the following need to be accommodated:

- Microbial kinetics – describes the rate at which the *combination* of microbe, substrate, and electron acceptor operates. This is intrinsically site specific, applies to microbe scales, and is dependent on the composting time course.
- Electron acceptor distribution.
- Any relevant substrate variation.

The electron acceptor distribution and substrate variation extend microbial kinetics into three dimensional space and are necessary to describe the degradation of any system larger than microbe scale. This is a very good starting point for bottom-up modelling.

However, the level of complexity that ensues from applying this causation too literally becomes a modelling question needing resolution; a composting model which attempts to accommodate the different degradation rates between cell walls and cell contents will be overwhelmed with detail. Yet it would be unwise to ignore this variation totally.

This composting situation has many similarities to that being grappled with in the ecological discipline where Grimm et al. (2005) proposed the 'Medawar zone' as being that point in model complexity where the payoff is maximised; make the model too complex and the payoff is reduced. The payoff in their context was a combination of usefulness and realism.

Grimm et al. argued that pattern-oriented modelling can identify patterns that contain information on the internal organisation of the system. Indeed Grimm et al. wrote: *Patterns are defining characteristics of a system and often, therefore, indicators of* ***essential underlying processes and structures*** (my emphasis). Later they wrote: *Ideally, the patterns used to design a model occur at different spatial and temporal scales and different hierarchical levels, because the key to understanding complex systems often lies in understanding how processes on different scales and hierarchical levels are bound to each other.*

In the case of ecology these patterns have to be 'uncovered' and this differs from composting in that the dominant pattern in composting (electron acceptor distribution) is able to be **determined** by meshing diffusion laws with microbial kinetics in the context of the particle – the *essential underlying processes and structures* of Grimm et al. can be identified.

The formulation in Chapman (2008) was influenced by needing to know the odour production potential of compost; leading to the need to understand the distribution of **anaerobic** electron acceptors as these generate the most odorous compounds. This formulation 'happened' to result in a satisfyingly high regression with the data. However, such accidental ways of hitting the Medawar zone are suboptimal; consequently a more formal way for location of the Medawar zone is proposed below.

It is suggested that there are two tools that can be used to help identify the patterns that can lead to an optimum structure for our models. The notions of:

- Rootedness – the degree to which the model parameters incorporate the fundamental laws and processes as these are the mathematical version of the causation noted above.
- Seamlessness – A model which seamlessly flows from the fundamental laws and processes at the atomic scale to the macro-scale features of an operating compost pile would have the greatest potential understanding. Any compromises (either mathematical or derivational) that are necessary to find a solution will reduce seamlessness. Seamlessness is a thinking tool used

as a prompt for the necessity for maintaining the essential causation. It can help us locate the Medawar zone.

It is further suggested that a useful modelling consideration is to concentrate on the areas where seamlessness is compromised. These being areas where there is a *potential constraint* on our total understanding due to expression of the underlying causation being compromised. Information carrying tools can then be considered as vehicles to ‘carry’ the information across this constraint, thereby maintaining seamlessness in the model.

2.1 Rootedness

From the perspective of causation raised above, rootedness in composting relates to the microbe/substrate kinetic; for which microbial kinetic models (first-order kinetics, Monod kinetics etc) are the mathematical description of this process. However, it was argued in Chapman (2008) that the notion of a rate constant (a parameter in microbial kinetics formulations) should be based on its more fundamental units of electron acceptor and substrate (see Table 1). Electron acceptor and substrate are the two chemical compounds necessary to formulate a chemical equation of composting. This is a useful starting point for any analysis as it is also the point in the complexity where mass balance laws and thermodynamics meet chemistry and by its association with this chemistry, the microbial system. Other convenient relationships occur at this point, in particular the different forms of rate constant have a common chemistry and can be related to each other by stoichiometry – this includes the rate constant forms that are based on volatile solids.

Table 1 Microbial kinetic rate constants based on their fundamental stoichiometric unit (from Chapman, 2008).

		SUBSTRATE		
		Fast	*Slow*	*Humification*
ELECTRON ACCEPTOR	*Oxygen*	$k_{O2(f)}$	$k_{O2(s)}$	$k_{O2(h)}$
	Nitrate	$k_{anoxic(f)}$	$k_{anoxic\ (s)}$	$k_{anoxic\ (h)}$
	Anaerobic	$k_{an(f)}$	$k_{an(s)}$	$k_{an(h)}$

Substrate in this context may be a group of compounds with similar degradation characteristics and equally wide range of chemical formulations, so a **balanced** chemical equation would be a compromise. However, this is sufficiently fundamental for this author although others may wish to go deeper.

Basing a composting model on the chemical components of composting will *carry* these fundamental roots into subsequent calculations.

2.2 Seamlessness

Seamlessness relates to the degree in which the model roots (the substrates and electron acceptors noted in Table 1 in conjunction with the microbes that influence the value of the rate constants) are preserved in the model structure.

A model is deemed seamless if the roots, in particular the contributing fundamental laws and processes, either:

- have a place in the model, or
- their effect can be identified in the model output.

Seamlessness can be seen as a thread carrying causation from the mathematised roots in an unbroken path to the application of the model; a thinking tool whose *sole* task is to ensure that causation is represented in the model output. If seamlessness is compromised then the cause and effect linkage is also compromised, and this becomes a potential ‘constraint’ on the model’s effectiveness.

For example, if an electron acceptor is **not** evenly distributed throughout a system, then there should be some part (or parameter) in the model where any effect of this uneven distribution can influence the model output. If this is not present, then the model is, in effect, based on an ‘average’ (specifically the proportionate contribution) of all the participating electron acceptors. Much potentially useful detail will be missed by this structural oversight. This can include structural errors in determining parameter values as detailed in Chapman (2009); where for example, not accommodating this variability would mean the calculated rate constant is net of all contributing rate constants but represent none of them adequately, limiting comparisons between experiments. This rate constant error will in turn generate an error in determining substrate concentration (if this is being determined from the experimental data).

Preserving seamlessness requires continuous vigilance. One could build a seamless model in several ways:

- By considering *scale*, where the beginning scale for composting would the sub-particle (microbe) scale distribution of electron acceptors, then scaling up to a particle, a pile, and eventually a technology utilising the pile.
- By widening the *context,* for which constraints on seamlessness would emerge as the composting pile moved from the controlled laboratory conditions that are, for example implicit in the rate constant, to the real-world application of the technology where rate constants (along with other parameters) change with temperature – (Chapman, 2011).
- Or by choosing a convenient *analytical unit* and ensuring that seamlessness is maintained within this analytical boundary – (Chapman, 2010).

2.3 Constraint resolution

The notion of a constraint is intended to draw attention to the **potential** limits of the model’s effectiveness. Resolving all constraints will result in a high quality model as this will enable expression of *all* the fundamental laws and processes (the roots) in the model output.

For composting, maintaining seamlessness between the microbiological fundamentals (taken as being a specific electron acceptor/substrate combination in a microbial kinetics model - Table 1) and the pile, within the context of changing environmental factors, required three different organizational considerations to cope with the ensuing constraints (discussed in more detail in Chapman (2010)):

1. Where computational units with a physical manifestation occurred:
    a. the particle based, *analytical boundary* separating the air phase from the composting phase; and
    b. the notion of a sub-particle scale *micro-environment* arising from using a finite element method. In this case, a finite time element generated a finite volume when diffusion laws were incorporated which accommodated emerging substrate variability.
2. Where new or existing parameters adequately served the information carrying tasks. For example:

    a. VOR as a single parameter encapsulating all the complexity of microbial kinetics and carrying the result seamlessly across to diffusion laws.
    b. Z encapsulating the net effect of environmental, microbial, and diffusion law interactions. This enabled all these effects to be incorporated into the physically manifest micro-environment.
    c. Aerobic proportion (Φ) which links particle size and z with microbial kinetics and therefore links all the fundamentals to the observed composting rate of a particle.
    d. The notion of a ‘combined’ parameter, being one whose value encapsulates other fundamental parameters. The influence of these other aspects (which can include the model structure) is reflected in the value that is experimentally determined. For example, the effect of a diffusible substrate is incorporated into the rate constant if a non-diffusible substrate solution of diffusion law is used and the rate constant is determined from the experimental data (Chapman, 2009). Surface roughness is incorporated into the diffusion coefficient and E(0) when not entered as a separate parameter (Chapman, 2009).
3. Where mathematical procedures proved useful:
    a. Modelling constants that become variables when environmental conditions change could be retained as constants when using a finite element method.
    b. The use of overlaying, interdependent sets of equations and iteration as a way of solving the mathematical problem of over-parameterization. Only those parameters necessary for a particular set of equations need to be *computationally* visible.
    c. Boundary conditions providing mechanisms by which interactions between the computational units could be managed (e.g. Henry’s law managing the boundary conditions between the air phase and the particle analytical unit).

Of particular interest to this author is the role that a physical manifestation (micro-environments) can serve in maintaining seamlessness. Micro-environments are formed using the fundamental laws and processes for their **information carrying tasks**. By maintaining seamlessness in the formation of each micro-environment, a *structure* is formed which reflects the *essential underlying processes*. Any subsequent calculation which utilises the micro-environment will ‘carry’ this seamlessness (and the underlying fundamental laws and processes) into subsequent calculations (and also other *structures* that may be formed at higher levels in the *hierarchy*).

This is the computational power of structure – a digital-age abacus. Whether they are called micro-environments or patterns does not impact on the computational power of the structure.

3 Discussion

Grimm et al. (2005) noted that “....*the key to understanding complex systems often lies in understanding how processes on different scales and hierarchical levels are bound to each other”*. They also noted that “*No general framework for designing, testing, and analysing bottom-up models has yet been established.....*”

Although different from the ecological perspective of Grimm et al., composting is a complex microbial ecology for which the micro-environment model (detailed in Chapman (2008)) explained the composting data with an $r^2 > 0.99$ (Chapman, 2009) i.e. it produced a very good fit. This was despite a number of relevant, but clearly unnecessary details being deliberately excluded in the formulation. To achieve the high regression coefficient, the micro-environment model devised by

Chapman would appear to contain *sufficient* complexity to explain the underlying processes, but not so much that it suffers the over-parameterization problems experienced by Hamelers (2001).

It is suggested here that the reason for the high regression coefficient is that this solution generated a **pattern[1] that reflected the underlying processes and how they are bound together**. For example, the zero-order oxygen consumption kinetic enabled determination of an oxygen penetration distance. This simple (and arguably incorrect) formulation generated an adequate description of the actual transition from a microbial kinetic dominated by oxygen to a kinetic dominated by other electron acceptors (anoxic and anaerobic). The **pattern** arising from the different microbial kinetics was fully reflected in the formulation, albeit with some loss of precision in the location of the transition. That is, some of the detail may be imprecise, in particular a slightly different value for the parameters (Chapman, 2009), but the overall model precision is not compromised as the pattern is fully explained.

The success of the model was a surprise even to the author; a pragmatic model that was also precise seemed to be a contradiction, yet such a state is entirely consistent with the Medawar zone of Grimm et al. (2005). Which leads, if it is indeed in the Medawar zone, to the question as to how Chapman's solution arrived at this place?

In retrospect, it can be seen that the pragmatic perspective adopted at the formulation stage contributed greatly to locating the Medawar zone in composting. Of particular relevance was the starting question which was the need to have some measure of potential odour production, which leads directly to consideration of the proportion of a particle which is anaerobic. Anaerobic is a term applied to a type of electron acceptor and can only occur in that part which is not aerobic.

But the same 'Medawar zone' is also uncovered by maintaining seamlessness, as any particular rate constant used to describe microbial kinetics is specific to an electron acceptor, forcing consideration of electron acceptor distribution. However, the biological response to the physical conditions does not fit neatly into a mathematical solution no matter how rigorous the analysis (the transition from aerobic to anoxic is a fuzzy transition and not amenable to high precision). Consequently, the zero-order oxygen consumption kinetic was adequate for the task of identifying the transition, with the added advantage of seamlessly carrying the temperature effect on all of: rate constant, oxygen solubility in water and diffusion coefficient, into the formulation (Chapman, 2010).

Chapman's 'solution' at the particle scale necessitated resolution of **all** the constraints (of which the electron acceptor distribution was only one) that emerged as increasing layers of model complexity were placed over the fundamental microbial kinetics that generated the pattern. The number of constraints meant that it was only possible to resolve these constraints *as far as possible.* Compromises needed to be made at many points for which the 'best mix' of compromises became a consideration (a Medawar zone within a Medawar zone perhaps).

Resolving constraints can be aided by considering the relative importance of each of the elements that generate the patterns. Low importance details (such as the **precise** location of the electron acceptor transition) could be ignored with little impact on model precision; locating the Medawar zone would be enhanced by this. This is discussed further in Chapman (2010).

---

[1] Grimm et al. considers a pattern to be: *observations of any kind showing non-random structure and therefore containing information on the mechanisms from which they emerge. Complex systems contain patterns at different hierarchical levels and scales.*

4 Conclusion

A complex system is most easily comprehended in small pieces. Each piece is a part of the whole, but no piece fully explains the whole, yet each piece has a deterministic relationship to the others. It is suggested here that formulations which carry *all* of the information inherent in the complexity are possible, but these formulations can only be 'solved' for a small part of the total system complexity. Specifically, we can 'know' only a small part of the system at any one computation, but are not limited as to which part can be 'known'. Understanding complexity therefore involves questions of the starting assumptions; i.e. what do we want to know? The part of the system complexity which is most useful to the modeller can be chosen as a part of the starting assumptions.

A model's applicability will be enhanced if causation to the fundamental laws and processes that occur at the base of any complex system is maintained. Causation can be maintained by resolving any constraints that arise if seamlessness is compromised.

This causation can be 'carried' in emergent parameters that are derived using seamlessness. These embed considerable information into the value of a single parameter. Much of this information is stored in the context which lies behind the value of each contributing parameter (discussed more fully in Chapman (2011); consequently this information is held in non-modelling space, but remains associated with it. Changing the context then gives us a tool by which we can gain insights into the behaviour of the complexity. This change in context can be as narrow as changing temperatures to simulate the real world (as demonstrated in the Q_{10} example in Chapman, (2010)), to the effect of changes in the container design that houses a compost pile. Each change will have a deterministic effect on the value of the emergent parameters.

A further part of this embedded information characteristic is the formation of computational units. Computational units are parts of the system whose behaviour is sufficiently similar that they can be treated as a single unit for calculation purposes. In the case of composting, these computational units have a physical form (a volume and location in space in the case of Chapman's (2008) micro-environment). When formed using seamlessness, these computational units become carriers of substantial information, and can form the fundamental building blocks of pattern-oriented modelling. In a sense these forms could be considered 'agents' such as used in agent-based modelling, albeit not necessarily autonomous agents. The potential for understanding complexity is enhanced by this formulation. For example, the micro-environments (patterns) of Chapman: allow for determination of the composting rate by simple summation; can identify anoxic and anaerobic zones by difference; can change size by gaining or losing water; and the volume part of a micro-environment's formulation leads to the optimising parameter aerobic proportion. In this manner, the structure of the model replicates the structure of the underlying processes; composting is after all a sum of millions of micro-organisms growing as they can, given the resources available to them and the conditions they find themselves in; a micro-environment groups these micro-organisms into volumes of compost containing similar composting rates.

In the case of composting, computational units also identify a lower limit to the scale of the analysis necessary to generate an adequate description of the system; this lower limit arises from the time interval chosen for the analysis and the effect of this length of time on the micro-environment thickness. The substrate at this scale becomes a mix of cell walls and cell contents; the 'system' rate constant will be net of the wall and contents rate constants and their proportionate contributions. A reasonable stance as rate constants must be determined experimentally anyway.

These are useful attributes for building models.

The ecological notions of a Medawar zone and a system of hierarchies as proposed by Grimm et al. (2005) were found useful to reconcile the output (which indicated a high precision model), with the numerous compromises (many of which could appear to invalidate the model) needed to form these computational units. Such an apparent contradiction is however entirely consistent with the notion of a Medawar zone.

A number of thinking tools are proposed to assist in locating the optimum formulation for any particular model:

- Maintaining seamlessness ensures that the fundamental laws and processes play their due role in the formulation.
- A constraint occurred when seamlessness seems to be compromised. Resolving these constraints was useful for finding model areas requiring resolution. Constraints to seamlessness indicate potential limits to the model's application.
- The use of hierarchies had the effect of layering the complexity into easier to comprehend parts. These layers are also useful in the mathematical sphere as it enables the complexity to be separated into smaller computational chunks. It is argued that hierarchies become a useful holder of the context.
- Establishing a synergy between the patterns in nature and the patterns that the model generate enhances the model's ability to reflect the real-world.

5 Bibliography

Chapman, P. D. (2008). *Application of diffusion laws to composting: theory, implications, and experimental testing.* Christchurch: Unpublished doctoral thesis, Lincoln University: Downloadable version at: http//researcharchive.lincoln.ac.nz/dspace/handle/10182/819.

Chapman, P. D. (2010). *Modelling composting complexity: the use of emergent, information rich, computational units as a solution to the over parameterization problem.* Retrieved from paulchapman.nz: http://www.paulchapman.nz/papers/Modelling-composting-complexity.pdf

Chapman, P. D. (2009). *Parameter determination in composting - Part I: The use of overlaying, interdependent sets of equations as a solution to the over parameterization problem.* Retrieved from paulchapman.nz: http://www.paulchapman.nz/papers/parameter-determination-PartI.pdf

Chapman, P. D. (2009). *Parameter determination in composting - Part II: Incorporating substrate diffusion and substrate solubilisation.* Retrieved from paulchapman.nz: http://www.paulchapman.nz/papers/Parameter-determination-PartII.pdf

Chapman, P. D. (2009). *Parameter determination in composting - Part III: Analytical boundary.* Retrieved from paulchapman.nz: http://www.paulchapman.nz/papers/Parameter-determination-PartIII.pdf

Chapman, P. D. (2011). *The use of context and hierarchies to extend seamlessness into technology choice.* Retrieved from paulchapman.nz: http://www.paulchapman.nz/papers/context&hierarchies-to-extend-seamlessness-into-technologies.pdf

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., et al. (2005). Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science* , 987-991.

Hamelers, H. V. (2001). *A mathematical model for composting kinetics* - *Unpublished doctoral thesis.* Netherlands: Wageningen University.