Supplementary Information for
Information Theory: A Foundation for Complexity Science

Amos Golan and John Harte

Amos Golan
Email: agolan@american.edu

This PDF file includes:

Box 1. The Structure of DynaMETE

DynaMETE (1) is a theory of dynamic complex systems in which the state variables are changing over time. It combines mechanistic, bottom-up, modeling to describe dynamics at the micro scale with top-down MaxEnt to infer distributions over the micro scale variables. Instantaneous values of the macro scale state variables and their first time derivatives constrain the MaxEnt procedure, and at any moment in time, the transition functions incorporating mechanisms at the micro scale may depend on the macro scale state variables as well as on the micro scale variables.

We illustrate the approach here with an ecosystem toy model that is simpler than the full DynaMETE described in the main text. Consider a dynamic state variable, N, equal to the total number of trees in a forest, and a fixed number of species, S. In this toy model, the single micro scale variable, n, is the abundance of an arbitrary species, and its time derivative is given by the mechanistic transition function $f(n, N(t))$. The notation implies that micro scale dynamics may depend on the macro scale variable. For simplicity, the METE variables, E (total metabolic rate) and ε (an individual's metabolic rate) are ignored here because we are focusing on the basic structure of this approach to hybridizing mechanism with MaxEnt. The goal is to calculate the probability distribution $P(n, N)$ and the expected time evolution of N under a perturbation in the initial value of N or in the transition function, f.

Suppose at time t we know N, dN/dt, and the two Lagrange multipliers λ_1 and λ_2, which satisfy the constraints:

$$N(t) = S\sum_n nP(n, N(t), \lambda_1(t), \lambda_2(t)) \tag{S1}$$

$$\frac{dN(t)}{dt} = S\sum_n f\big(n, N(t)\big)P(n, N(t),\ \lambda_1(t), \lambda_2(t)) \tag{S2}$$

Under these constraints (see Box 1 in the main text)

$$P\big(n, N(t), \lambda_1(t),\ \lambda_2(t)\big) = \frac{e^{-\lambda_1(t)n}e^{-\lambda_2(t)f(n,N(t))}}{z(t)} \tag{S3}$$

To update N, dN/dt, λ_1 and λ_2 we first update N by a unit time step: $N(t+1) = N(t) +\ dN/dt$, and update $f(n, N)$ by substitution.

Then we update dN/dt:

$$\frac{dN(t+1)}{Sdt} = \sum_n f\big(n, N(t+1)\big)P(n, N(t+1),\ \lambda_1(t), \lambda_2(t) \tag{S4}$$

Now using the constraints imposed by $N(t+1)$, $dN(t+1)/dt$, $f(n, N(t+1))$, we can update the λ_i (again, see Box 1 in the main text). This iterative procedure readily generalizes to multiple macro- and micro-scale variables as shown in (1).

An Example: Population Dynamics in a Disturbed Multi-Species Ecosystem

Ecologists typically model the density-dependent growth of a population using the logistic equation:

$$dn/dt = b_0 n - d_0 n^2 \tag{S5}$$

where the birth rate is linear in the population size, n, while the death rate is quadratic, indicating that crowding increases the per-capita death rate.

Suppose, however, that within a community of populations of multiple species, any one species feels the effect of crowding both from individuals, n, within that species, and also from the summed population of all the species, N. N may vary in time as the individual populations of the species rise and fall over time. We modify Eq. S5 as:

$$dn/dt = b_0 n - \ d_0 n^2 N/N_0 \tag{S6}$$

where the right hand side of Eq. S6 is the transition function $f(n, N(t))$ appearing in Eqs. S2 – S4, and N_0 is a constant which we take to be the initial steady-state, value of N.

In the initial steady state, the mean of the transition function over the distribution P is zero, the mean of n is N_0/S_0, and $P = Z^{-1}e^{-\lambda_1 n}$. Using the iteration process described above, the time trajectories of N and the Lagrange multipliers, $\lambda_i(t)$, following a perturbation either in the demographic rate constants or in the initial value of N can be computed.

Stochasticity is easily introduced into the transition function, f, as is extinction or migration leading to changes in the value of S. For more realism, the state variable E (the community metabolic rate), and the corresponding micro-level variable ε, (the metabolic rate of an individual), can be included (1).

Although a full analysis of the behavior of even just the simplified dynamical system (Eqs. S1-S6) is beyond the scope of this Perspective, preliminary analysis indicates multiple steady states, differing in the form of $P(n)$, accessible from other steady states by different types of initial perturbations in rate constants such as d_0 or displacements of N, and exhibiting hysteresis when a rate constant is first increased, say, and then decreased to its original value. Thus examination of the shape of abundance distributions in ecology may be able to provide insight into the perturbation history of the system. Further analysis of the potentially rich array of behaviors of hybrid mechanism-plus-MaxEnt dynamics is underway.

Box 2. A Conditional Markov Model

This Box complements the written description in the text.

Notations. The observed information in state j ($j = 1, \dots, K$) of individual i ($i = 1, \dots, n$) in period t ($t = 1, \dots, T$) is $y_{i,t,j}$. For each individual i, $y_{i,t,j} = 1$ if state j is observed at period t, and $y_{i,t,j} = 0$ for all other $K - 1$ states. The observed states are connected to the individual's unobserved probability $q_{i,t,j}$ in the following way: $y_{i,t,j} = q_{i,t,j} + \varepsilon_{i,t,j}$ where $\varepsilon_{i,t,j}$ captures the uncertainty.

***Object of interest*.** The $K \times K$ matrix of transition probabilities p_{kj} ($k, j = 1, \dots, K$), representing the probability of transitioning from state k to state j within time period t to $t + 1$. It relates to the unobserved probability $q_{i,t,j}$ via $q_{i,t+1,j} = \sum_{k=1}^{K} p_{kj} y_{i,t,k}$ where $\sum_j p_{kj} = 1$.

The Basic Model

In terms of the observed states, the Markov model is

$$y_{i,t+1,j} = \sum_{k=1}^{K} p_{kj} y_{i,t,k} + \varepsilon_{i,t,j} \tag{S7}$$

where the uncertainty $\varepsilon_{i,t,j}$ is now expressed as uncertainty on the model as a whole, such as possible misspecification or ambiguity. The goal of the MaxEnt approach is to is maximize the entropy of the P's subject to the above constraints (Eq. S1) and the K normalizations. This model, however, is too simple: it does not take advantage of all the available information.

The Realistic and Generalized Model

***Introducing Covariates and Accommodating for Model Ambiguity*.** Let X be L exogenous and environmental variables with elements $x_{i,t,l}$ for $l = 1, \dots, L$. The variables may be entity dependent or independent. The relationship between the observed states $y_{i,t,j}$, the unknown probabilities P, and the exogenous information X is

$$\sum_{t=2}^{T} \sum_{i=1}^{n} y_{itj} x_{itl} = \sum_{t=1}^{T-1} \sum_{i=1}^{n} \sum_{k=1}^{K} p_{kj} y_{itk} x_{itl} + \sum_{t=1}^{T-1} \sum_{i=1}^{n} \varepsilon_{itj} x_{itl}.$$

These are just the product moments once we introduced the X's into Eq. S1.

We are interested in both P and ε without imposing more structures. Therefore, the problem is underdetermined, so we follow the logic of the MaxEnt. But first, we need to convert the ε's to probabilities (Step 1).

Step 1: Converting Uncertainty to Probability Distributions. Think of the ε's as expected values of a discrete random variable V with corresponding probabilities W, so $\varepsilon_{itj} = \sum_m w_{itjm} v_{tjm}$ where $\sum_m w_{itjm} = 1$, $M \geq 2$, and $v_{tjm} \in [-1, 1]$. For example, if $M = 3$, $v_{tj} = (-1, 0, 1)$ for all t and j.

Step 2: The Updated Constraints. We now express the constraints in terms of V and W:

$$\sum_{t=2}^{T}\sum_{i=1}^{n} y_{itj}x_{itl} = \sum_{t=1}^{T-1}\sum_{i=1}^{n}\sum_{k=1}^{K} p_{kj}y_{itk}x_{itl} + \sum_{t=1}^{T-1}\sum_{i=1}^{n} \varepsilon_{itj}x_{itl}$$
$$= \sum_{t=1}^{T-1}\sum_{i=1}^{n}\sum_{k=1}^{K} p_{kj}y_{itk}x_{itl} + \sum_{t=1}^{T-1}\sum_{i=1}^{n}\sum_{m=1}^{M} w_{itjm}v_{tjm}x_{itl}\,. \tag{S8}$$

Finally, the information-theoretic goal is to maximize the Shannon entropy of P and W subject to constraints (Eq. 2) and normalizations for P and W. The inferred transitions are

$$p_{kj}^{*} = \frac{\exp\left(-\sum_{t=1}^{T-1}\sum_{il} y_{itk}x_{itl}\lambda_{jl}^{*}\right)}{\sum_{j} exp\left(-\sum_{t=1}^{T-1}\sum_{il} y_{itk}x_{itl}\lambda_{jl}^{*}\right)} \equiv \frac{\exp\left(-\sum_{t=1}^{T-1}\sum_{il} y_{itk}x_{itl}\lambda_{jl}^{*}\right)}{\mathrm{Z}_k(\lambda^{*})}$$

where $\{\lambda_{jl}^{*}\}$ is the set of inferred Lagrange multipliers associated with the $K \times L$ constraints (Eq. S2). Similarly, the inferred uncertainty is $\varepsilon_{itk}^{*} = \sum_m w_{itjm}^{*}v_{tjm}$ where

$$w_{itjm}^{*} = \frac{exp\left(-\sum_l x_{itl}v_{tjm}\lambda_{jl}^{*}\right)}{\sum_m exp\left(-\sum_l x_{itl}v_{tjm}\lambda_{jl}^{*}\right)} = \frac{exp\left(-\sum_l x_{itl}v_{tjm}\lambda_{jl}^{*}\right)}{\Psi_{itk}(\lambda^{*})}.$$

Though at a first glance it may seem that this is a complicated constrained optimization problem with many parameters, in fact it is not. The real parameters here are the Lagrange multipliers. The number of those is not changed with the generalization done here. Using duality theory – the principle that any optimization problem can be specified in two different ways, say one constrained with respect to the probabilities P's and W's, and one unconstrained with respect to the Lagrange multipliers – this can be easily observed via the unconstrained model (which is a function of the λ's):

$$\ell(\lambda) = \sum_{t=2}^{T}\sum_{ijl} y_{itj}x_{itl}\lambda_{jl} + \sum_{k} log[\mathrm{Z}_k(\lambda)] \sum_{itj} log\left[\Psi_{itj}(\lambda)\right].$$

Technically, $\ell(\lambda)$ can be interpreted as a generalized likelihood function (2, 3).

***Inferred Causal Influence*.** The marginal effects of X on P capture the direct effect of a small change in X on P while holding everything else fixed. If X_l is exogenous (it is some observed information that is determined outside the system; the value of this variable is independent of the states of the other variables in the system), then the change in p_{kj} as a result of a change in that variable is the inferred causal influence of that variable, such as the effect of global warming or a certain policy, on P. Mathematically, these causal effects can be calculated for both continuous and discrete causes.

SI References

1. J. Harte, K. Umemura, M. Brush, DynaMETE: a hybrid MaxEnt-plus-mechanism theory of dynamic macroecology. *Ecology Letters* **24**, 935–949 (2021).
2. A. Golan, *Foundations of Info-Metrics: Modeling, Inference, and Imperfect Information* (Oxford University Press, 2018).
3. A. Golan, G. Judge, J. Perloff, A Generalized Maximum Entropy Approach to Recovering Information from Multinomial Response Data. *J Am Stat Assoc* **91**, 841–853 (1996).