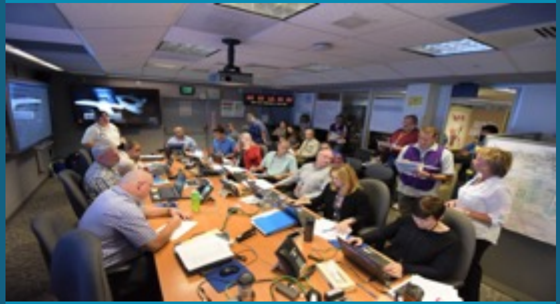
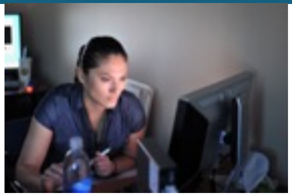
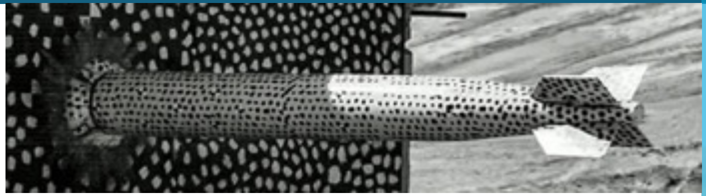
Sandia National Laboratories

Detecting physical pathways with software profiling







ESCO 2022

Andrew Steyer (Center for Computing Research, Sandia National Laboratories)

with Luca Bertagna, Graham Harper, Jerry Watkins (Sandia National Laboratories)

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

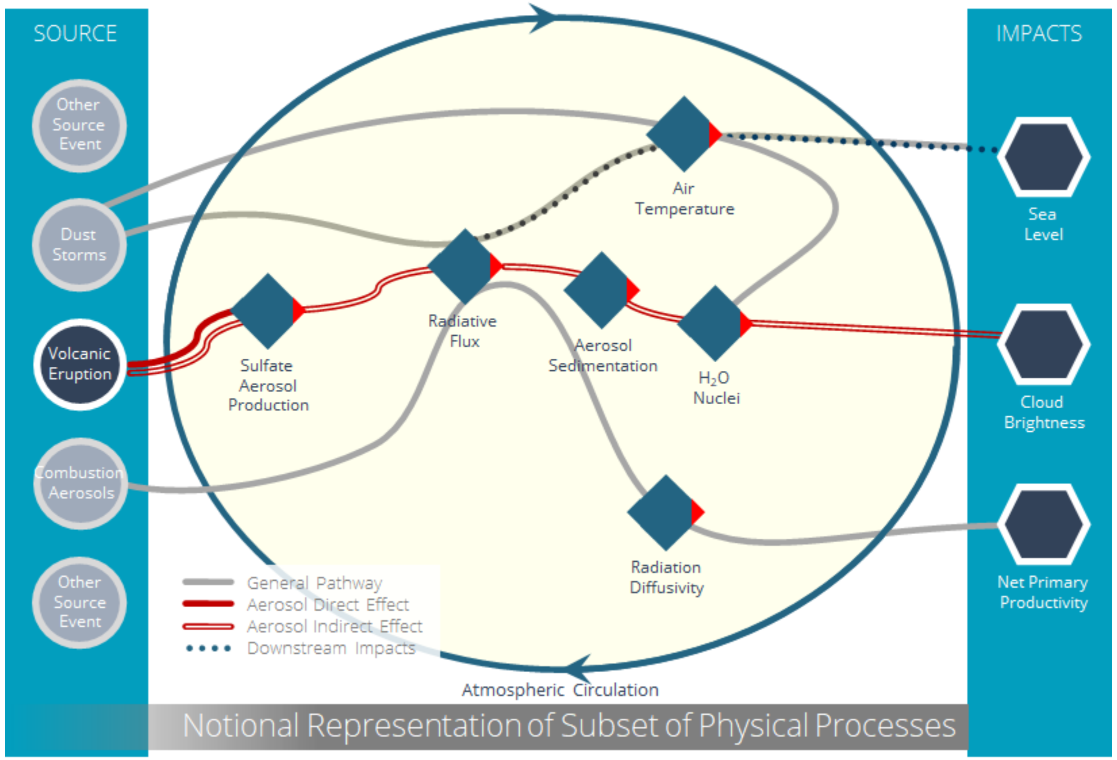
SAND2022-7528 C

E3SM model and CLDERA

E3SM
Energy Exascale
Earth System Model

E3SM is a fully coupled (atmosphere, ocean, land, ice, …) Earth system model developed and funded by the US Department of Energy (DOE) with focus on water cycle, biogeochemistry and energy, and the cryosphere.

Open source!: https://github.com/E3SM-Project/E3SM
V2 released September 2021!



https://education.nationalgeographic.org/resource/calderas: "A caldera is a large depression formed when a volcano erupts and collapses"

CLDERA -- CLimate impact: Determining Etiology thRough pAthways

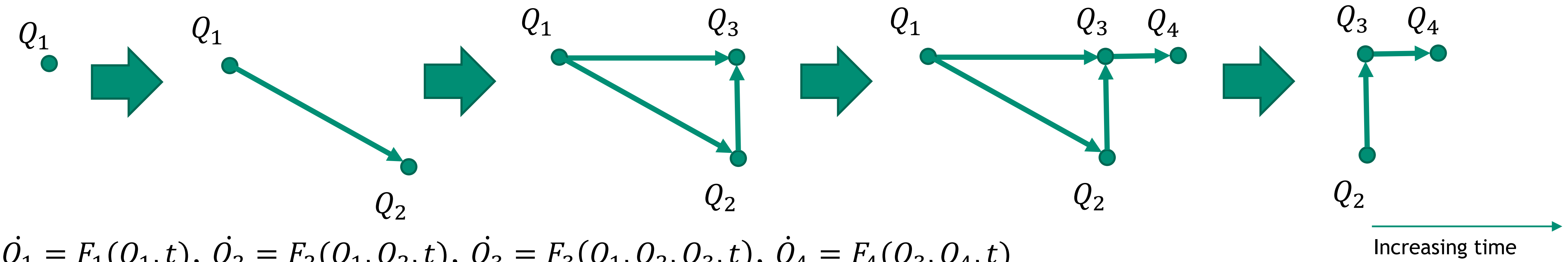
Compute physical pathways due to sources (focus on exemplar Pinatubo eruption) and attribute climate impacts caused by this source.

Simulated pathways thrust: to develop a new data-science informed modeling framework to dynamically trace and rank connective relationships ("pathways") between a climate source and its impact(s) using E3SM simulations.

Pathways and networks

Heuristic: A *pathway* is the chain of physical processes from source to impacts and their evolution in space and time.

Mathematically, this can (and has) been interpreted as a (time-dependent) graph/network representation of the dynamics of some system (typically a differential equation or a discrete/algebraic relative).



$$\dot{Q}_1 = F_1(Q_1, t),\ \dot{Q}_2 = F_2(Q_1, Q_2, t),\ \dot{Q}_3 = F_3(Q_1, Q_2, Q_3, t),\ \dot{Q}_4 = F_4(Q_3, Q_4, t)$$

Large body of literature on network dynamics, dynamic networks, and related areas going back many years: (Barzel and Barabási, 2013), (Anderson et al., 2020), (Glass et al., 2021), (Feinberg and Horn, 1973), (Holme and Saramäki, 2012), (Nicosia, V. et al. 2012).

Arise in social science, chemistry, biology, engineering, infectious disease modeling, etc - however most work focuses on pre-existing networks and/or static networks.

Our problem – given a source term or forcing of a complex system (such as climate modeled by E3SM), what is the dynamic network structure describing/approximating the forced dynamics i.e. what dynamic network is activated by the forcing?

Pathways from forcings in E3SM

E3SM: At each model step $m = 0,1,2,\ldots$ solve for $u = u(x, t_m)$ (with initial value $\mathrm{u}_0(x) = u(x, t_0)$):

$$F(u_t, u, x, t, \rho; V) = 0, \qquad \rho \text{ - model parameters}, \qquad V = V(x,t) \text{ - external forcing (e.g. volcano)}$$

Set of quantities of interest (QoI) - these might or might NOT (e.g. global mean temp.) be state variables:

$Q = \{Q_{\alpha_j} = \mathrm{Q}_{\alpha_j}(u, x, t, \rho; V) : j = 1, \ldots, N_Q\}$ - the α_j's can be strings e.g. SO2.

$\bar{Q}_{\alpha_j}(u, x, t, \rho) = Q_{\alpha_j}(u, x, t, \rho; 0)$ - the "unforced" QoI's.

Pinatubo – The external forcing V is a temporary source term for SO2 concentration $Q_{CSO2} = Q_{CSO2}(u, x, t, \rho)$:

$\frac{\partial Q_{CSO2}}{\partial t} + \vec{u} \cdot \nabla Q_{CSO2} = V + S$, where $V(\cdot, t) = 0$ when $t \notin [t_0, t_V]$,

$S = S(u, x, t, \rho)$ - the usual source/sink terms, $\vec{u} = \vec{u}(x,t)$ - velocity field from atmosphere.

All other variables not directly dependent on V: $Q_{\alpha_j} = Q_{\alpha_j}(u, x, t, \rho)$ when $\alpha_j \neq CSO2$.

Pathways are determined by constructing **directed** acyclic graphs (DAGs) related to the downstream effects of V on other QoIs.

Rigorously defining pathways

Given a set of QoIs $\mathcal{Q}$, a *pathway of* $\mathcal{Q}$ is a set $\mathcal{G} = \{\mathcal{G}_n\}_{n=0}^{N}$ where each $\mathcal{G}_n$ is a DAG with nodes representing elements of $\mathcal{Q}$ and where an arrow from node q_1 to q_2 means that the QoI represented by q_2 is impacted (according to certain criteria) by the QoI represented by q_1 at the time t_n.

"Certain criteria" – some way to quantify when a QoI is anomalous (we are being intentionally vague!).

Pathways are determined by the structure of the E3SM software, which limits how QoIs can interact.

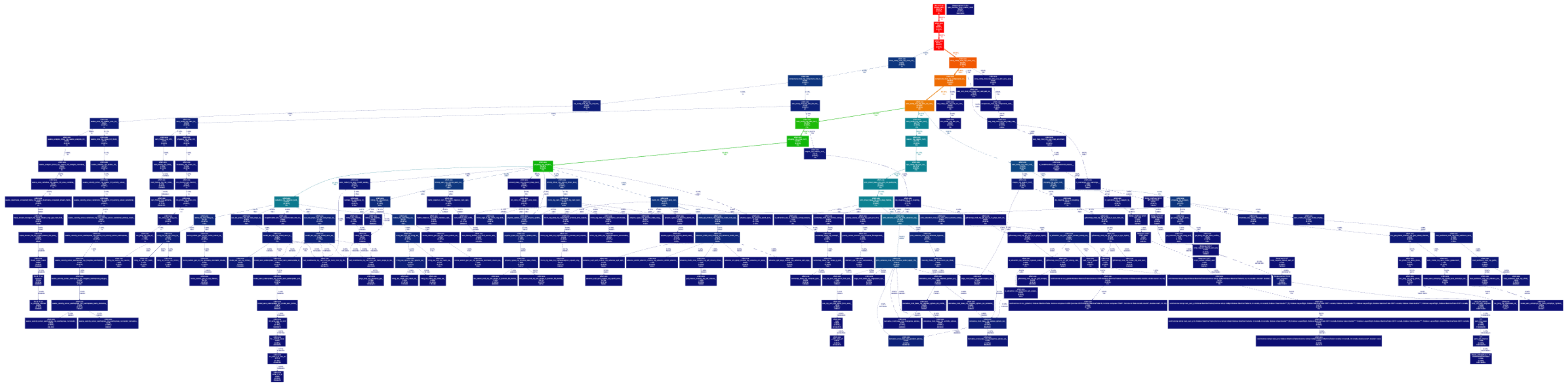
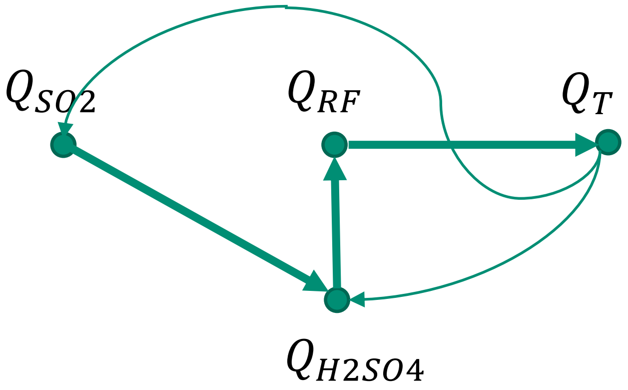


Figure: Partial subroutine-dependency graph E3SM (atmosphere dynamic core and physics.

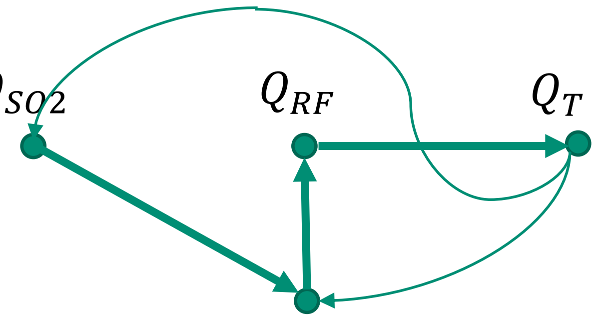
Consider the following example:

Q_{SO2} - global SO2
Q_{H2SO4} - global H2SO4
Q_{RF} - global radiative forcing
Q_T - global mean temperature



Thin lines denote indirect feedbacks – increase Temperature can over time affect SO2 and H2SO4 concentrations through various bio-chemical mechanisms, but not as strongly as volcanic eruption. Would like to ignore relatively weak indirect effects as an approximation to avoid using cyclic graphs.

SO2 pathway assumptions



Express the pathway QoIs as a function of the forcing and the unforced QoIs:

$$\begin{aligned} Q_{SO2} &= \bar{Q}_{SO2} + \phi_{SO2}(V) + \varepsilon_{SO2} \\ Q_{H2SO4} &= \bar{Q}_{H2SO4} + \phi_{H2SO4}(\Delta Q_{SO2}) + \varepsilon_{H2SO4} \\ Q_{RF} &= \bar{Q}_{RF} + \phi_{RF}(\Delta Q_{H2SO4}) + \varepsilon_{RF} \\ Q_T &= \bar{Q}_T + \phi_T(\Delta Q_{RF}) + \varepsilon_T \end{aligned}$$

$\Delta Q_\alpha := Q_\alpha - \bar{Q}_\alpha$ “anomaly in Q_α”

$\varepsilon_\alpha = \varepsilon_\alpha(u, x, t, \rho; V)$

In practice, the ΔQ’s are primarily dependent on the volcanic emissions: $||\varepsilon_*|| \ll ||\phi_*||$.

Open question: Can this be made rigorous? What should be the assumptions on the structure of the governing equations and pathway graph(s), the QoIs, and the forcing V, imply that $||\varepsilon_*|| \ll ||\phi_*||$.

Finding the right path

Given a set of QoIs $\mathcal{Q} = \{Q_{\alpha_j}\}_{n=1}^{N_Q}$ there are $G(N_Q)$ possible DAGs with nodes $\mathcal{Q}$, $G(k) = \sum_{j=1}^{k} (-1)^{j+1} \binom{k}{j} 2^{j(k-j)} G(k-j)$.

$\mathcal{D}$ - the set of all possible DAGs with N_Q vertices (many will be nonphysical e.g. surface air pressure does not depend directly on ocean salinity).

The E3SM software *restricts* pathways by enforcing physically realistic assumptions i.e. there exists $\mathcal{G}_{E3SM} \in \mathcal{D}$ so that any computed pathway $\mathcal{G} = \{\mathcal{G}_n\}_{n=0}^{N}$ of $\mathcal{Q}$ is such that $\mathcal{G}_n$ is a sub-DAG of $\mathcal{G}_{E3SM}$. Intuitively, $\mathcal{G}_{E3SM}$ is the ambient space of potential pathways.

To restrict the search space to quantities with the most impact, we introduce a base DAG $\mathcal{G}_{Base}$ which is a sub-DAG of $\mathcal{G}_{E3SM}$ and search for pathways within this $\mathcal{G}_{Base}$.

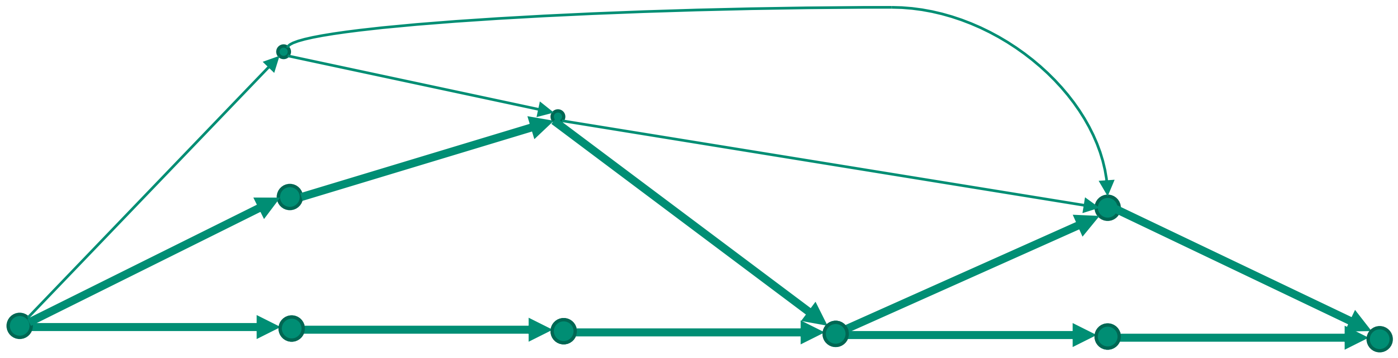


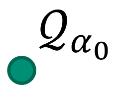
Figure (left). The full graph represents $\mathcal{G}_{E3SM}$, while the subgraph consisting of edges with wide lines and nodes with large circles represents $\mathcal{G}_{Base}$

Algorithm for discovering pathways

Problem: compute pathway $\mathcal{G} = \{\mathcal{G}_n\}_{n=0}^{N}$ from inputs $\mathcal{G}_{Base}$, $\mathcal{Q} = \{Q_{\alpha_j}\}_{n=1}^{N_Q}$, forcing V.

For simplicity, assume Q_{α_1} is the only directly forced QoI e.g. single species aerosol injection.

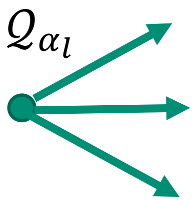
Let $\mathcal{M}(Q_{\alpha_j}) := \{Q_{\alpha_{j,m_1}}, \ldots, Q_{\alpha_{j,m_j}}\} \in \mathcal{Q}$ denote the children of Q_{α_j} in $\mathcal{G}_{Base}$. Let $\mathcal{G}_0$ be the single node graph



Let $\mathcal{M}o \leftarrow \{Q_{\alpha_1}\}$ (initially monitor Q_{α_1})

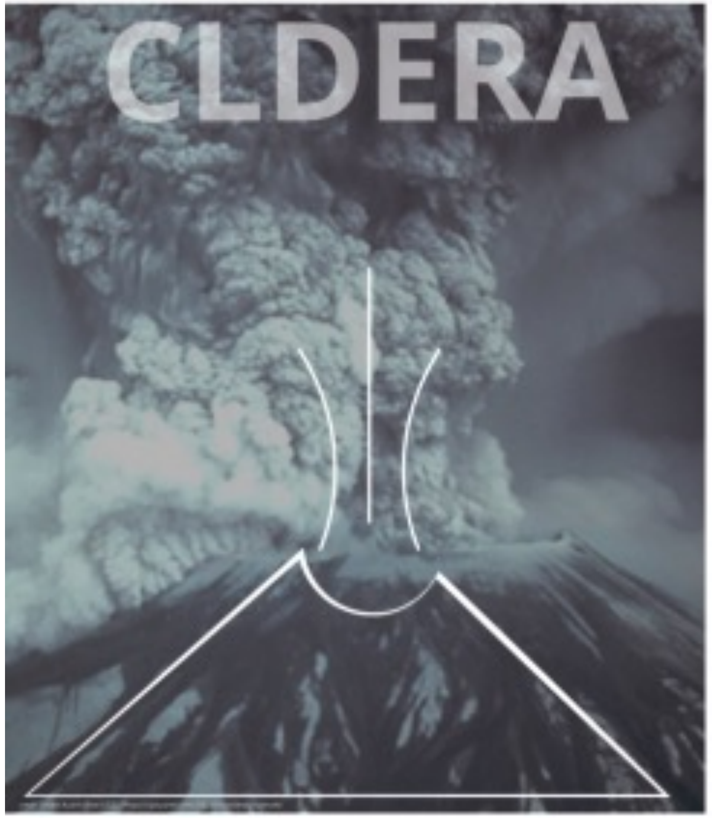
For $n = 1,2, \ldots, N$
$\mathcal{G}_n \leftarrow \mathcal{G}_{n-1}$
For all l such that $Q_{\alpha_l} \in \mathcal{M}o$
For all $j = 1, \ldots, m_l$
If $Q_{\alpha_{l,j}}$ fails some test, append

to $\mathcal{G}_n$ and $\mathcal{M}o \leftarrow \mathcal{M}o \cup \mathcal{M}(Q_{\alpha_{l,j}})$.
End For
If Q_{α_l} passes some test, then remove



from $\mathcal{G}_n$ (may continue to monitor Q_{α_l} if desired)
End For
End For

Simulating Pinatubo



See (Ramachandran et al. 2000) and (Stenchikov et al. 1998) for references on Pinatubo effects. We concentrate on changes to aerosol optical depth in the visual spectrum (mean temperature effects take longer times to show up).

(1) 10Tg of SO2 gas is injected into the atmosphere

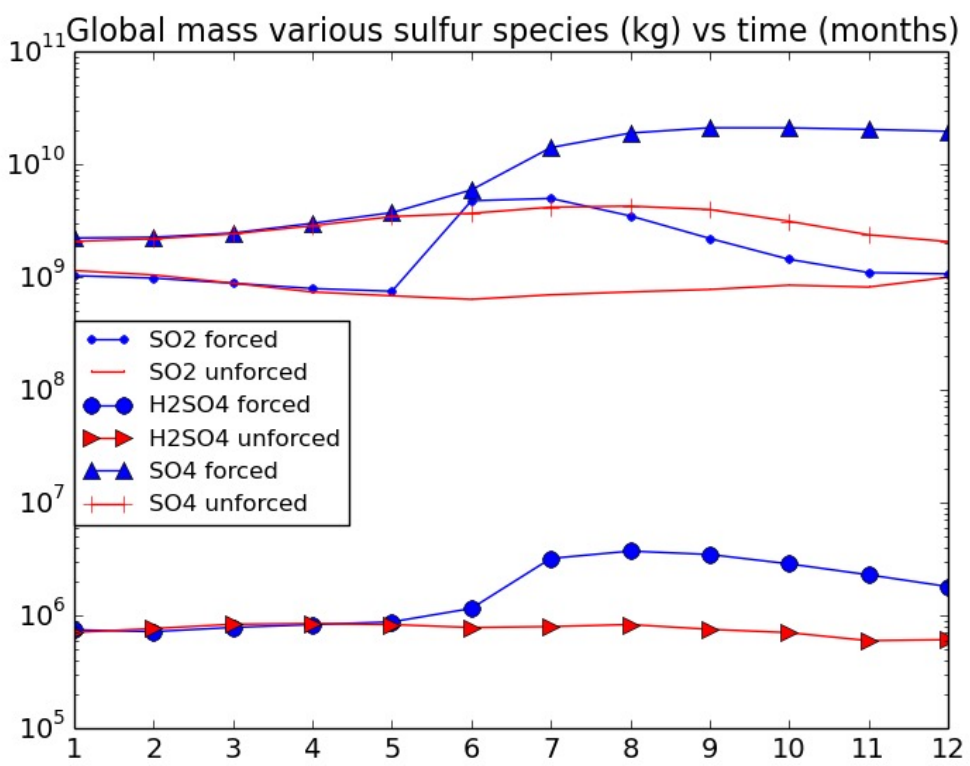
(2) SO2 reacts with OH to make H2SO4 and with O2 to make SO4.

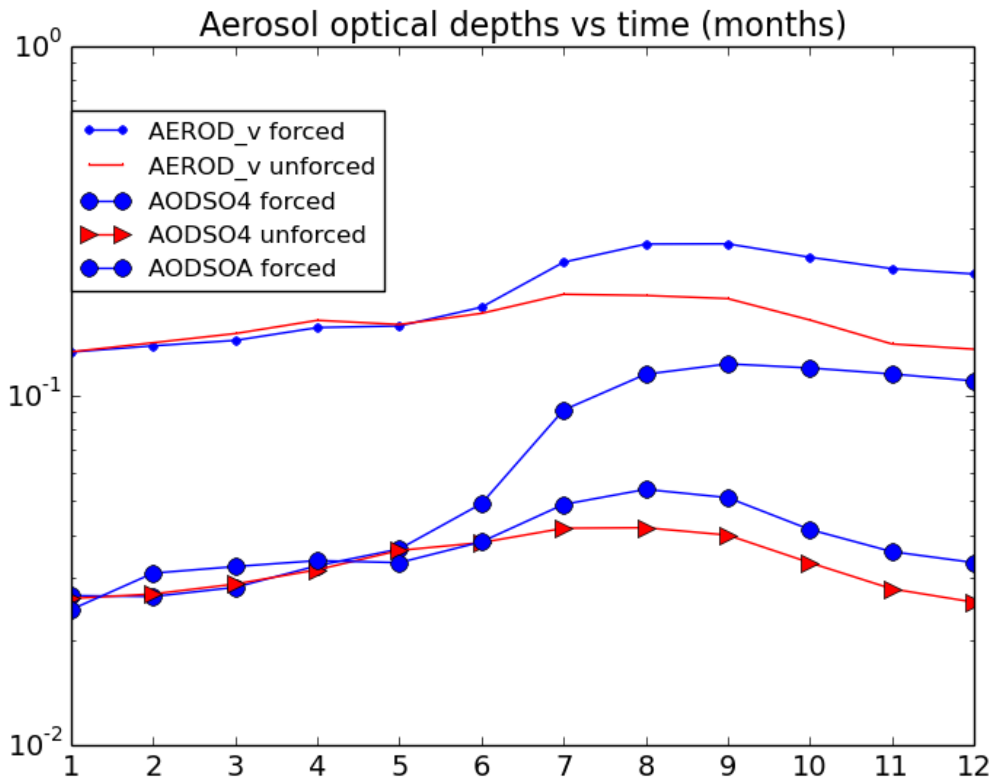
(3) SO4 and H2SO4 increase the aerosol optical depth (AOD) of the atmosphere.

(4) SO4 and H2SO4 slowly rain and sediment out of the atmosphere and AOD reduces to normal levels.

Results are for a "low-res" 1-year simulation: an ne30pg2 horizontal grid is used - about 80km, 1° resolution for dynamics (velocity, temperature,etc); 40km ,0.5^0 for physics (chemistry, aerosol transport,etc) with an 1800s time-step (300s sub-stepping for dynamics).

Mount Pinatubo pathway (graphs from output)

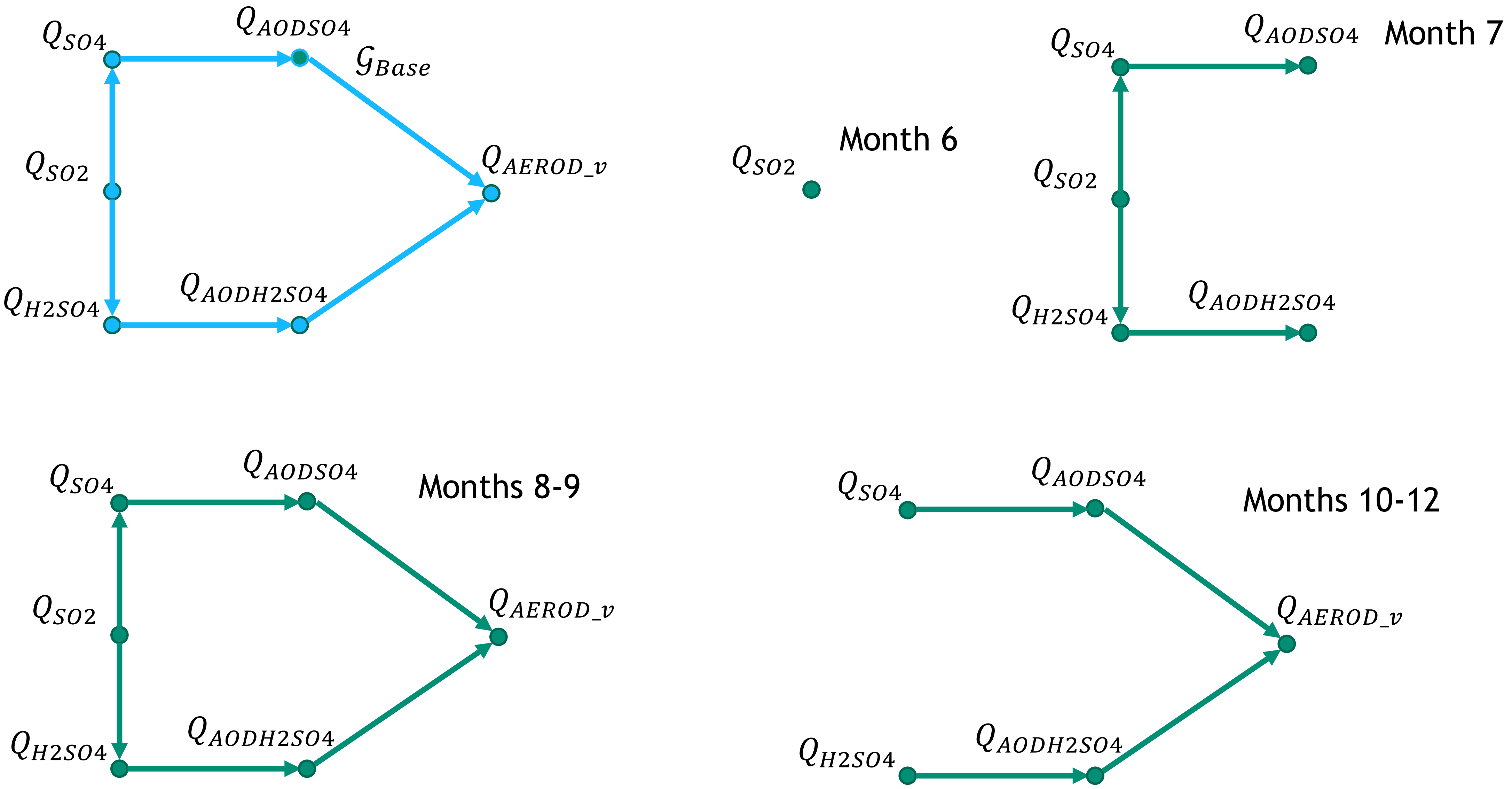




Time series of monthly means for (left) various sulfur species and (right) various aerosol optical depths of a 1 year low resolution E3SM simulation of a forced run with Mount Pinatubo injection 10Tg of SO2 in to the atmosphere and an unforced run with no volcanic eruption and nearby Initial condition.

Pinatubo Pathway (DAGs)

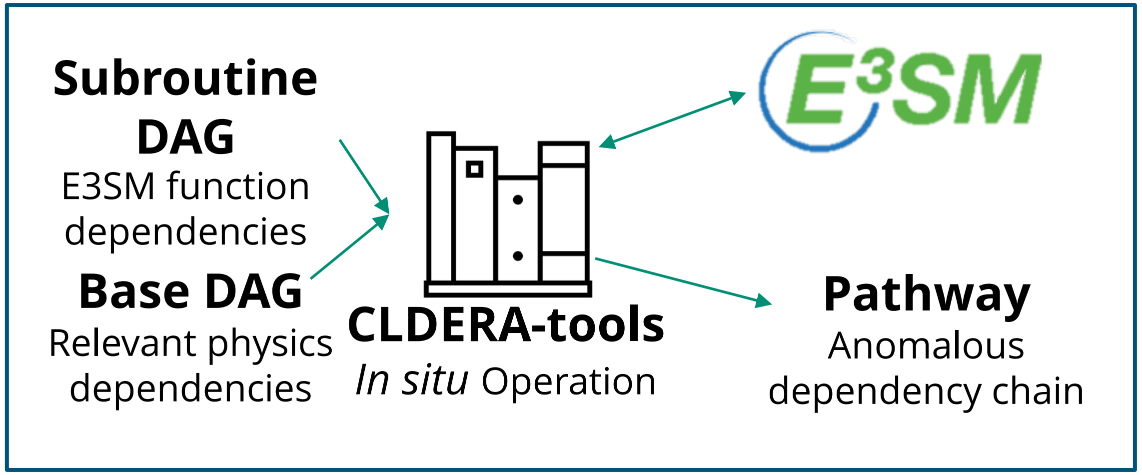
Criteria for adding/removing nodes: add/remove node when QoI is more/less than double the unforced mean.



CLDERA-tools

Very preferable to avoid computing DAGs by hand from output!

Software interface (currently in testing) called CLDERA-tools for in-situ discovery of pathways.
Inputs: $\mathcal{Q}$, $\mathcal{G}_{Base}$
Outputs: Some pathway $\mathcal{G}$.



In-situ Pathways: Instrument E3SM and develop Directed Acyclic Graphs (DAGs) of process interactions executed within code. Minimally invasive – only requires an interface to graph QoIs from E3SM and runs concurrently.

Investigate: detect if QoI is anomalous, when to add/remove nodes and edges, fast/slow processes, including spatial dependencies, …

References

Anderson, D., Brunner, J., Craciun, G., and Johnston, M. On classes of reaction networks and their associated polynomial dynamical systems. *J Math Chem* **58,** 1895–1925 (2020) https://doi.org/10.1007/s10910-020-01148-9.

Barzel, B., Barabási, AL. Universality in network dynamics. *Nature Phys* **9,** 673–681 (2013) https://doi.org/10.1038/nphys2741.

Glass, D.S., Jin, X. & Riedel-Kruse, I.H. Nonlinear delay differential equations and their application to modeling biological network motifs. *Nat Commun* **12,** 1788 (2021). https://doi.org/10.1038/s41467-021-21700-8.

Golaz, J.-C. et al. The DOE E3SM Model Version 2: Overview of the physical model, Earch and Space Science Open Archive, pp. 61, (2022) https://doi.org/10.1002/essoar.10511174.1

Holme, P. and Saramäki, J. Temporal Networks, *Physics Reports,* **519**, 97-124 (2012) https://doi.org/10.1016/j.physrep.2012.03.001.

Nicosia, V., Tang, J., Musolesi, M., Russo, G., Mascolo, C., and Latora, V. Components in time-varying graphs, *Chaos*, **22**, 023101 (2012) https://doi.org/10.1063/1.3697996.

Ramachandran, S., Ramaswamy, V., Stenchikov, A., Robock A. Radiative impact of the Mount Pinatubo volcanic eruption: Lower stratospheric response, *Climate and Dynamics*, **105**, 24409-24429 (2000) https://doi.org/10.1029/2000JD900355.

Stenchikov, G. et al. Radiative forcing from the 1991 Mount Pinatubo volcanic eruption, *Journal of Geophysical Research Atmospheres*, **103**, 13837-13857 (1998) https://doi.org/10.1029/98JD00693.