Supporting Information. Grace, J. B., and K. M. Irvine. 2020. Scientist's guide to developing explanatory statistical models using causal analysis principles. Ecology.

Appendix S3
Wildfire Structural Equation Model Analysis

Part S1: Background information on applications of structural equation modeling (SEM).

There are a number of modern treatments of SEM (Hoyle 2012, Hancock and Mueller 2013, Kline 2016), some directed at natural scientists (Grace et al. 2012, 2015, Shipley 2016); thus, extensive coverage of the subject is not given here. The first-generation implementation of SEM was little more than a loose confederation of (a) path analysis, implemented using standard regression techniques, (b) factor analysis, which required maximum likelihood for estimation, and (c) simultaneous equations, which are a central preoccupation of econometricians. The second-generation implementation was marked by a revolutionary advance based on the analysis of covariance structures (Jöreskog 1972). This implementation, which remains a useful and commonly applied approach today, is characterized by a global-solution process that tests the covariance relationships implied by a model against those found in the data (e.g., the 'lavaan' R package, Rosseel 2012). A third generation of SEM is currently unfolding, inspired by Pearl's broad vision for what SEM can be as a companion to causal analysis (outlined in Grace et al. 2012 and Chen, Pearl and Kline 2019). This third-generation implementation aspires to accommodate the latest statistical methods of estimation and modeling. This requires moving beyond the constraints of covariance structure analysis and allowing for locally-determined specifications. Pioneering work by Shipley (2000a, 2000b, 2009) in the development of methods for testing d-separation claims and now the development of software to convert lists of classical statistical models into networks (Lefcheck 2016) opens up the range of possible techniques while also breaking down the artificial barrier between statistical traditions. Bayesian implementations of SE models are becoming common place (Arhonditsis 2006, Lee 2007, Lee and Song 2012, Gimenez et al. 2012, Souchay et al. 2018) as are a number of other fusions (e.g., phylogenetic SEM - Gonzalez-Voyer and von Hardenberg 2014, spatially-explicit SEM – Lamb et al. 2014). A large number of SEM training modules can be found at https://bit.ly/graceSEM.

References

Arhonditsis, G. B., C. A. Stow, L. J. Steinberg, M. A. Kenney, R. C. Lathrop, S. J. McBride, and K. H. Reckhow. 2006. Exploring ecological patterns with structural equation modeling and Bayesian analysis. Ecological Modelling 192:385–409.

Chen, B., J. Pearl, and R. Kline, R. 2019. Graphical tools for linear path models. Psychometrika (in press).

Gimenez, O., T. Anker-Nilssen, and V. Grosbois. 2012. Exploring causal pathways in demographic parameter variation: path analysis of mark–recapture data. Methods in Ecology and Evolution 3:427–432.

Gonzalez-Voyer, A. and A. Von Hardenberg. 2014. An introduction to phylogenetic path analysis. Pages 201-229 *in* L. Z. Garamszegi, editor, Modern phylogenetic comparative methods and their application in evolutionary biology. Springer, Berlin, Heidelberg.

Hancock, G. R. and R. O. Mueller, eds. 2013. Structural equation modeling: A second course. Information Age Publishing, Charlotte, North Carolina, USA.

Hoyle, R. H., editor, 2012. Handbook of structural equation modeling. Guilford press, New York, New York, USA.

Jöreskog, K. G. 1973. A general method for estimating a linear structural equation system. Pages 85–112 in A. S. Goldberger and O. D. Duncan, editors. Structural equation models in the social sciences. Seminar Press, New York, New York, USA.

Kline, R. B. 2016. Principles and Practice of Structural Equation Modeling. Fourth edition. Guilford Press, New York, New York, USA.

Lamb, E. G., K. L. Mengersen, K. J. Stewart, U. Attanayake, and S. D. Siciliano. 2014. Spatially explicit structural equation modeling. Ecology 95:2434-2442.

Lee, S. Y. 2007. Structural equation modeling: a Bayesian approach. John Wiley & Sons, New York, New York, USA.

Lee, S. Y., and X. Y. Song. 2012. Basic and advanced Bayesian structural equation modeling: With applications in the medical and behavioral sciences. John Wiley & Sons, New York, New York, USA.

Lefcheck, J. S. 2016. piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. Methods in Ecology and Evolution 7:573-579.

Rosseel, Y. 2012. Lavaan: An R package for structural equation modeling and more. Journal of Statistical Software 48:1-36.

Shipley, B. 2000a. Cause and correlation in biology: A user's guide to path analysis, structural equations, and causal inference. Cambridge University Press, Cambridge.

Shipley, B. 2000b. A new inferential test for path models based on directed acyclic graphs. Structural Equation Modeling 7:206-218.

Shipley, B. 2009. Confirmatory path analysis in a generalized multilevel context. Ecology 90:363–368.

Shipley, B. 2016. Cause and correlation in biology. Second edition, Cambridge University Press, Cambridge, UK.

Souchay, G., R. E. van Wijk, M. Schaub, and S. Bauer. 2018. Identifying drivers of breeding success in a long-distance migrant using structural equation modelling. Oikos 127:125-133.

Part S2: Code (in Box S1) and results (in Boxes S2-S4) presented in Figure 4 of main text. Data for this illustration is in “Data S1”. A separate R code file can be found in “Data S3”. Analyses were run using R version 3.5.3 (R Core Team 2019).

Box S1. Code for estimation and evaluation of wildfire recovery models in Figure 4.

```
#############################################
##### SEM Analyses using 'piecewiseSEM' #####
#############################################
library(piecewiseSEM)

fdat <- read.csv("AppendixS1_data.csv")

# model 1
pw.mod1 <- psem(lm(vegcover ~ firesev + age + elev, data=fdat),
                lm(elev ~ coastdist, data=fdat),
                lm(age ~ coastdist + elev, data=fdat),
                lm(firesev ~ age + elev, data=fdat))

# model 2 – alt. to model 1 with link from E -> F omitted
pw.mod2 <- psem(lm(vegcover ~ firesev + age + elev, data=fdat),
                lm(elev ~ coastdist, data=fdat),
                lm(age ~ coastdist + elev, data=fdat),
                lm(firesev ~ age, data=fdat))

# model 3 – alt. to model 2 with link from C -> F added
pw.mod3 <- psem(lm(vegcover ~ firesev + age + elev, data=fdat),
                lm(elev ~ coastdist, data=fdat),
                lm(age ~ coastdist + elev, data=fdat),
                lm(firesev ~ age + coastdist, data=fdat))

# model 4 – alt. to model 1 with additional link from C -> V
pw.mod4 <- psem(lm(vegcover ~ firesev + age + elev + coastdist,
                   data=fdat),
                lm(elev ~ coastdist, data=fdat),
                lm(age ~ coastdist + elev, data=fdat),
                lm(firesev ~ age + elev, data=fdat))

# model 5 - simplification of model 1, omitting A -> V
pw.mod5 <- psem(lm(vegcover ~ firesev + elev, data=fdat),
                lm(elev ~ coastdist, data=fdat),
                lm(age ~ coastdist + elev, data=fdat),
                lm(firesev ~ age + elev, data=fdat))

# model 6 - simplification of model 1, omitting E -> V
pw.mod6 <- psem(lm(vegcover ~ firesev + age, data=fdat),
                lm(elev ~ coastdist, data=fdat),
                lm(age ~ coastdist + elev, data=fdat),
                lm(firesev ~ age + elev, data=fdat))
```

Box S2. Results: d-separation tests of independence claims for wildfire recovery models.

```
> dSep(pw.mod1)
              Independ.Claim Test.Type DF Crit.Value P.Value
1 vegcover ~ coastdist + ...      coef 85 -0.4317232 0.6670
2  firesev ~ coastdist + ...      coef 86 -0.7338175 0.4650

> dSep(pw.mod2)
              Independ.Claim Test.Type DF Crit.Value P.Value
1 vegcover ~ coastdist + ...      coef 85 -0.4317232 0.6670
2  firesev ~ coastdist + ...      coef 87 -1.6792289 0.0966
3       firesev ~ elev + ...      coef 86 -0.7933242 0.4297

> dSep(pw.mod3)
              Independ.Claim Test.Type DF Crit.Value P.Value
1 vegcover ~ coastdist + ...      coef 85 -0.4317232 0.6670
2       firesev ~ elev + ...      coef 86 -0.7933242 0.4297

> dSep(pw.mod4)
              Independ.Claim Test.Type DF Crit.Value P.Value
1  firesev ~ coastdist + ...      coef 86 -0.7338175 0.4650

> dSep(pw.mod5)
              Independ.Claim Test.Type DF Crit.Value P.Value
1 vegcover ~ coastdist + ...      coef 86  0.3469151 0.7295
2  firesev ~ coastdist + ...      coef 86 -0.7338175 0.4650
3       age ~ vegcover + ...      coef 85 -2.1711359 0.0327

> dSep(pw.mod6)
              Independ.Claim Test.Type DF Crit.Value P.Value
1  firesev ~ coastdist + ...      coef 86 -0.7338175 0.4650
2 vegcover ~ coastdist + ...      coef 86  1.0381223 0.3021
3      vegcover ~ elev + ...      coef 85  1.8984586 0.0610
```

Box S3. Model comparisons.

```
> ### Model comparison
> AIC(pw.mod1, aicc=T)
[1] 42.338  # lowest AICc
> AIC(pw.mod2, aicc=T)
[1] 45.208
> AIC(pw.mod3, aicc=T)
[1] 42.533
> AIC(pw.mod4, aicc=T)
[1] 44.414
> AIC(pw.mod5, aicc=T)
[1] 47.435
> AIC(pw.mod6, aicc=T)
[1] 48.062
```

Box S4. Coefficients from selected model.

```
> coefs(pw.mod1)
  Response Predictor Estimate Std.Error DF Crit.Value P.Value Std.Estimate
1 vegcover   firesev  -0.0594    0.0203 86    -2.9236  0.0044  -0.3095  **
2 vegcover       age  -0.0578    0.0267 86    -2.1674  0.0330  -0.2289   *
3 vegcover      elev   0.0249    0.0116 86     2.1381  0.0353   0.2026   *
4     elev coastdist  17.7205    2.4808 88     7.1430  0.0000   0.6058 ***
5      age coastdist  -7.5268    1.7307 87    -4.3490  0.0000  -0.5288 ***
6      age      elev   0.2013    0.0592 87     3.4028  0.0010   0.4138  **
7  firesev       age   0.6166    0.1241 87     4.9681  0.0000   0.4689 ***
8  firesev      elev  -0.1031    0.0604 87    -1.7071  0.0914  -0.1611

> rsquared(pw.mod1)
  Response   family     link method R.squared
1 vegcover gaussian identity   none 0.2595463
2     elev gaussian identity   none 0.3670083
3      age gaussian identity   none 0.1857412
4  firesev gaussian identity   none 0.2317276
```

Literature Cited

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.