Supporting Information

Wells et al. (Individual and temporal variation in pathogen load predicts long-term impacts of an emerging infectious disease)

Detailed description of the modelling framework

We implemented a stochastic individual-based simulation model of coupled Tasmanian devil demography and devil facial tumour disease (DFDT) epidemiology, for which we provide an overview of design, concept, and details (Grimm et al. 2006).

Purpose

The purpose of this model is to simulate the impact of DFTD on Tasmanian devil populations and validate model scenarios of different input parameters (26 model parameters assumed to be unknown and difficult or impossible to estimate from empirical studies, see **Table S1**) by matching known system level properties (disease prevalence and population structure) derived from a wild population studied over ten years after the emergence of DFTD (Hamede et al. 2015). In particular, running model scenarios for 100 years prior to, and after the introduction of DFTD, we explored the extent to which DFTD causes devil populations to decline or become extinct. Moreover, we aimed to explore whether input parameters such as the latency period of DFTD or the disease transmission frequencies among individuals of different ages can be identified by matching simulation scenarios to field patterns of devil demography and disease prevalence.

Entities and state variables

Entities in the model are individuals that move in weekly time steps (movement distance θ) within their home ranges and may potentially engage in disease-transmitting biting behaviour with other individuals (**Fig. 1**). Free-ranging individuals (i.e. those recruited into the population after 34 weeks of weaning in pouch and den) are characterized by seven state variables: sex (sex_i), age ($age_{i,t}$), home range centre (hr_{Xi} |hr_{Yi}), actual location (loc_{Xi} |loc_{Yi}), time of last reproduction event ($repT_{i,t}$), time of infection with DFTD ($infT_{i,t}$), and tumour load ($V_{i,t}$). The time of the last female reproduction events informs about the number of young recruited into the population (conditional that the respective females and their offspring survive the 34 weeks of pouch and den weaning time). The environment is a homogenous 15 × 15 km space, in which home range centres of individuals are randomly

located (coordinates of home range centres were drawn as random coordinates from a uniform distribution within the boundaries of the space).

Process overview and scheduling

In each time step (weekly), processes are scheduled in the following order: 1) reproduction of adult females and males (if the week matches the reproductive season), 2) recruitment of juveniles into the population, 3) natural death (independent of DFTD), 4) physical interaction and potential disease transmission, 5) growth of tumours, 6) DFTD-induced death, 7) movement of individuals, 8) aging of individuals. In the beginning of each time step, the week is attributed as the corresponding calendar week (with all simulations starting at the first week of a year) to account for seasonality in mating and reproduction.

Birth-death processes are modelled as probabilities according to specified input parameter values for each scenario. To avoid unrealistically large population sizes N_t, we assumed mortality rates for all age classes to gradually increase above a carrying capacity of $C = 300$ (using the function $min(0.1, \kappa + (1- \kappa) (0.1 / (1 + exp(-0.01(N_{t-1}- C)))))$, in which weekly mortality rates κ increases towards the maximum values of 0.1 if $N_{t-1} > C$; all finite population size estimates in our case study were below this value(Wells et al. 2017)). The carrying capacity of 300 individuals in an area of 225 km^2, or a density of 1.3 km^{-2}, is a typical density of disease-free devil populations (McCallum et al. 2009).

The force of infection $\lambda_{i,t}$, i.e. the probability that a susceptible individual i acquires DFTD at time t is given as the sum of the probabilities to have DFTD transmitted from any interacting infected individual k (with $k \in 1 \ldots K$, with K being the number of all individuals in the population excluding i):

$$\lambda_{i,t} = \sum_k \beta I_{k,t},$$

Here, β is the disease transmission coefficient and $I_{k,t}$ indicates infectious individuals k. For the transmission of DTFD, we may expect that the transmission of tumours depend on tumour size and also age and reproductive status, since we expect mature and reproductively active individuals to more often engage in aggressive interactions that may facilitate disease transmission by increased biting activity. To account for these possibilities we extended the basic model for $\lambda_{i,t}$ as follows:

$$\lambda_{\mathrm{i,t}} = \left[\sum_{k \in \mathrm{K}} \beta_{A(i)} \beta_{A(k)} \left(\frac{N_t}{C}\right)^{\delta} \left(\frac{1}{1+(1-r_{i,t})\omega}\right) \left(\frac{1}{1+(1-r_{k,t})\omega}\right) \left(\frac{V_{k,t}}{V_{max}}\right)^{\gamma}\right] I_{\eta}$$

Here, the disease transmission coefficient is composed of two factors $\beta_{A(i)}$ and $\beta_{A(k)}$, each of which accounts for the age-specific interaction and disease transmission rate for individual i

and k according to their age classes. N_t is the population size at time t; the scaling factor δ accounts for possible increase in interactions frequency with increasing population size if $\delta > 0$. The parameter $r_{i,t}$ is a Boolean indicator of whether an individual recently reproduced and ω is a scaling factor that determines the difference in $\lambda_{i,t}$ resulting from interactions of reproductively active and non-reproducing individuals. $V_{k,t}$ is the tumour load of individual k, V_{max} is the maximum tumour load, and γ is a scaling factor of how $\lambda_{i,t}$ changes with tumour load of infected individuals. The parameter I_η is a Boolean indicator of whether two individuals are located in a spatial distance $< \eta$ that allows interaction and disease transmission (i.e. only individuals in distances $< \eta$ can infect each other). We considered individuals as 'reproductively active' ($r_{i,t}$=1) for eight weeks after a reproduction event.

DFTD-induced mortality Ω_{size} account for tumour size with tumour size classes (< 50 cm^3, 50 – 100 cm^3, > 100 cm^3); the magnitudes of Ω_{size} as input parameters (i.e. changing 'virulence') were specified according to the uncertainty of estimates from field data (Wells et al. 2017).

Tumour growth was modelled as a logistic function with the growth parameter α sampled as an input parameter and maximum tumour load set to M_{max} = 202 cm^2 according to the maximum/asymptotic tumour mass reported elsewhere (Wells et al. 2017). We allowed for latency periods τ between infection and the onset of tumour growth, which was also sampled as an input parameter. We assumed no recovery from DFTD, which appears be very rare in the field (Pye et al. 2016). We did not explore the effects of repeated/secondary re-introduction of DFTD into the modelled population. While there are currently no quantitative information on the spatial spread and metapopulation dynamics of DFTD available, the relatively slow geographical spread of DFTD across Tasmania (two decades after emergence, some devil populations are still disease free within the ca. 68 km^2 sized landscape) suggest a minor role of re-introductions.

The model is stochastic in that demographic rates (survival and reproduction), disease-induced death and movement distances for each individual and week are drawn from random distributions, assuming Gaussian error of 10% of respective input parameters.

Emergence

Transmission of DFTD results from the interaction of susceptible and infected individuals, depending on how likely individuals are to interact according to movement patterns, the modelled interaction frequencies and disease progression/ tumour load. Population

fluctuations and changes in the prevalence of DFTD over time are emerging properties that result from the coupled demographic and epidemiological dynamics.

Design concepts. The seasonal demographic birth-death process and tumour growth in the model follows empirical analysis of devil survival and fecundity and disease progression from a case study (Wells et al. 2017). Movement of individuals follow expert knowledge (Hamede et al. 2009) and unpublished studies on the movement behaviour of devils by David Hamilton and Sebastien Comte (University of Tasmania). DFTD is assumed to be transmitted during physical interaction that involves biting tumour-carrying individuals.

Simulation scenarios. We drew 10^6 scenarios of random input parameter values using latin hypercube sampling (Stein 1981) to cover a large range of possible parameter combinations. All parameters were sampled from uniform distributions with ranges specified in **Table S1**. Notably, sampled scaling factor values of zero for δ, ω, and γ correspond to model scenarios with homogeneous interactions frequency and disease transmission rates independent of population size, reproductive status and tumour load, respectively, while values of η = 21 km assume that individuals can infect each other independent of spatial proximity (i.e. individuals across the entire study area can infect each other). We included in the sampled parameter space scenarios that excluded *i)* effects of tumour load on infection and survival propensity, *ii)* effect of spatial proximity on the force of infection between pairs of individuals and *iii)* both effects of tumour load and spatial proximity in each of 1,000 scenarios. This sampling design was used to explicitly assess the importance of modelling individual tumour load and space use for accurately representing the system dynamics.

For each scenario, we first ran 100 years (520 weeks) of simulations for disease free populations. For those scenarios in which devil population were stable (i.e. population always > 100 and < 400 individuals) and juveniles never comprised > 50% of the population (Wells et al. 2017), we introduced DFTD to a random selection of 5% of adult individuals (tumour sizes of randomly sampled uniformly of sizes 0.01 cm^3 to M_{max}) and then ran another 100 years of simulations. In total, devil populations were stable according to criteria specified above in 13,523 out of 10^6 scenarios (1.35% only).

Computation. We ran the model in R version 3.4.3 (R Development Core Team 2017). Computation of 10^6 scenarios on 500 parallel nodes on a computer cluster with Xeon CPU X5650 processor (2.67GHz) took ca. 200 hours.

Table S1. Parameter definitions and their values/ sampled ranges for the individual-based model of demographic and epidemiological dynamics of Tasmanian devils and devil facial tumour disease. Parameters sampled with changing values (drawn from uniform distributions) in each scenario are marked with an asterix. The parameter ranges are largely defined based on previous analysis of field data (Wells et al. 2017) unless defined otherwise in the detailed model descriptions.

Parameter	Symbol	Range / Unit	Description
Demography			
Carrying capacity	C	300	Carrying capacity of the modelled populations above which mortality increase with density.
Survival rates*	Φ_{age}	Φ_{PY}: 0.7 – 0.9 Φ_{DY}: 0.7 – 0.9 Φ_{Juv}: 0.4 – 0.9 Φ_{SA}: 0.4 – 0.9 Φ_{Y1}: 0.4 – 0.9 Φ_{Y2}: 0.4 – 0.9 Φ_{Y3}: 0.4 – 0.9 Φ_{Y5}: 0.1 – 0.7	Annual survival rates for different age groups (weeks of age range given in parenthesis). Φ_{PY}: pouch young (1 - 17); Φ_{DY}: young in den (18 - 34); Φ_{Juv}: juveniles (35-45); SA: subadults (46-51); Φ_{Y1}: one. year old (52-103); Φ_{Y2}: two years old (104-155); Φ_{Y3}: three-four years old (156-259); Φ_{Y5}: five years old and older (≥260).
Maximum age of survival	$MaxAge_{\Phi}$	7 years	Maximum age devils can reach in the model; individuals reaching this age will die.
Reproduction rates (females) *	Ψ_{age}	Ψ_{Y1}: 0 – 0.2 Ψ_{Y2}: 0.5 – 0.99 Ψ_{Y3}: 0.5 – 0.99	Probability of a mature female reproducing during the annual breeding season depending on age (weeks of age range given in parenthesis). Ψ_{Y1}: one year old (52-103); Ψ_{Y2}: two years old (104-155); Ψ_{Y3}: three-five years old (156-260).
Maximum age of reproduction	$MaxAge_{\Psi}$	5 years	Maximum age female devils can reproduce.
Reproduction season	$Week_{Repro}$	Calendar week 8 / 17 (min/max)	Calendar weeks in which female devils may reproduce.
Reproduction interval		45 weeks	Minimum time period between two reproductive events of females.
Movement distance	θ	2 – 10 km	Movement distance individuals move away from their home range centre in weekly time steps.
Epidemiology			
Minimum tumour load	m_0	0.0001 cm^2	Minimum tumour volume at onset of growth.
Maximum tumour load	M_{max}	202 cm^2	Maximum/ asymptotic tumour volume (as used in logistic growth curve).
Scale parameter of the logistic growth curve	α	0.02 – 0.1	Scale parameter of the logistic growth curve of tumours, given as a value of weekly growth.

Latency period	τ	0 – 150	Latency period between infection and onset of tumour growth in weeks.
DFTD-induced decrease in survival rates	Ω_{size}	Ω_1: 0.3 – 0.9 Ω_2: 0.1 – 0.6 Ω_3: 0.05 – 0.4	Odds ratios of the decrease in survival of individuals with certain tumour loads compared to uninfected individuals. Tumour load categories are Ω_1: < 50 cm^3; Ω_2: 50 – 100 cm^3; Ω_3: > 100 cm^3.
Interaction and disease transmission coefficients	β_{age}	β_{Juv}: 0 – 1 β_{Y1}: 0 – 1 β_{Y2}: 0 – 1 β_{Y3}: 0 – 1 β_{Y5}: 0 – 1	Interaction and disease transmission coefficients based on the interaction of susceptible and infectious individuals. β_{Juv}: juveniles (35-45); SA: subadults (46-51); β_{Y1}: one year old (52-103); β_{Y2}: two years old (104-155); β_{Y3}: three-four years old (156-259); β_{Y5}: five years old and older (≥260).
Tumour load – dependence in infection rates	γ	0 – 3	Scaling factor, which determines increases in infection rates with increasing tumour load of infected individuals.
Density - dependence of interaction frequency	δ	0 – 3	Scaling factor of possible increase in interactions frequency with increasing population size if $\delta > 0$.
Reproduction - dependence increase in disease transmission	ω	0 – 4	Scaling factor, which determines the difference in infection rates resulting from interactions between reproductively active and non-reproducing individuals.
Interaction distance for disease transmission	η	1 – 21 km	Distance over which individuals at different locations may interact and eventually transmit disease.

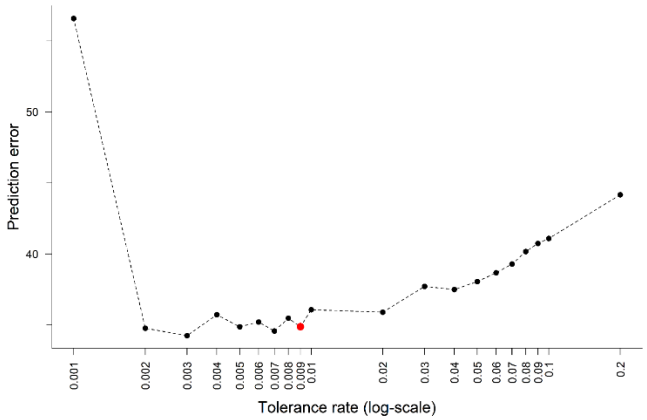

Figure S1. Prediction error for different tolerance rates (subset selection of simulation scenarios) for predicting the summary statistics from our case study, using the leave-one-out cross validation procedure of the 'cv4abc' function in the R package *abc*. Prediction errors are calculated using the neural network regression method, sample sizes of n=100 as samples sizes for the cross-validation samples and 'mode' as measures of central tendency from posterior distributions.

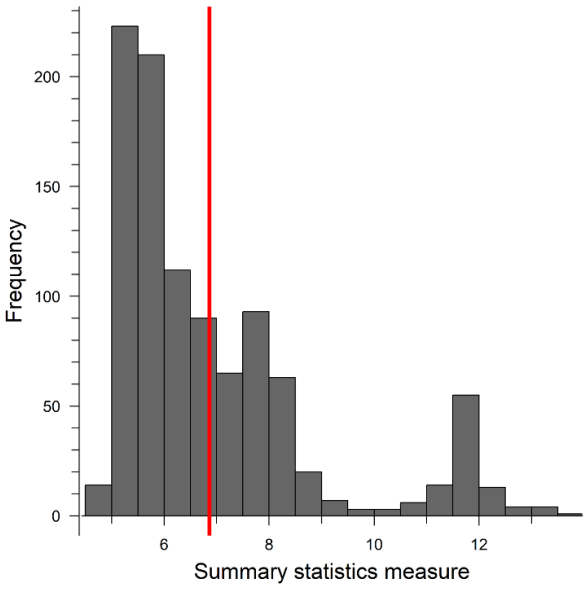

Figure S2. Frequency distribution of summary statistic measures (summed over all different summary statistic variables). The red bar shows the value for the summary statistics from our case study. Summary and goodness-of-fit test were performed with the 'gfit' function of the R package *abc*.

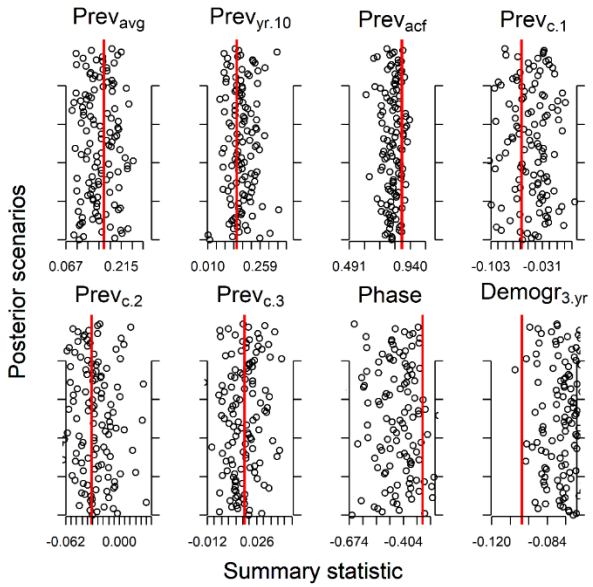

Figure S3. Summary statistic values for the empirical case study (red vertical lines) and each of the selected posterior scenarios (points). The horizontal distances between red lines and points indicate the differences in summary statistics between those derived from field data and those from selected scenarios. Summary statistics are given as **$Prev_{avg}$**: mean DFTD prevalence over the course of 10 years; **$Prev_{yr10}$**: mean DFTD prevalence in the 10^{th} year only; **$Prev_{acf}$**: autocorrelation value for prevalence values lagged over one time step; **$Prev_{c.1}$**, **$Prev_{c.2}$**, **$Prev\_{c.3}$**: three coefficients estimates of cubic regression model of the smoothed ordered difference in DFTD prevalence (fitting 3^{rd} order orthogonal polynomials of time for smoothed prevalence values using the loess function in R with degree of smoothing set to $\alpha = 0.75$); **Phase**: phase in seasonal population fluctuations as determined by a sinusoidal model fit to the trappable population over the course of ten years; **$Demogr_{3.yr}$**: regression coefficient of a linear model of the changing proportions of individuals ≥ 3years old in the trappable population over the course of ten years.

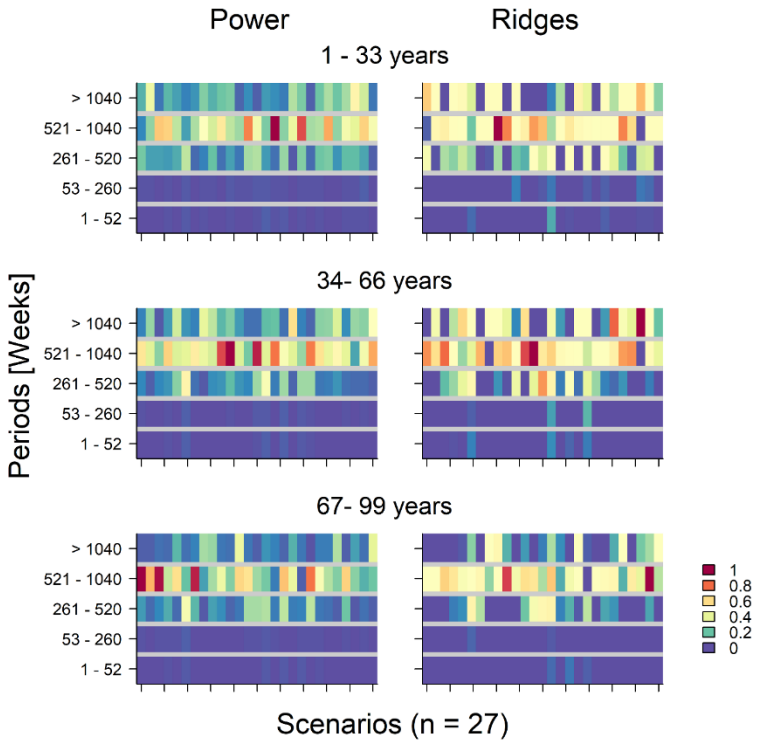

Figure S4. Relative intensities of wavelet power spectra and the presence of ridges (continuous high intensity of periodicity over time) in different periods and timespans after the introduction of devil facial tumour disease (DFTD) in those scenarios in which DFTD persisted in the population for at least 100 years of simulations (n = 61 out of 121 posterior samples). The different scenarios are plotted as small bars in arbitrary order along the x-axis. Colour spectra from blue to red indicated increasing relative spectral intensities in the different regions of the individual wavelet spectra and the relative presence of ridges (proportion of regions covered by ridges), respectively. Binning of wavelet spectra into regions was done over the five weekly time periods 1-52, 53-260, 261-520, 521-1040, and > 1040 weeks, respectively, and the three timespans 1-33, 34-66, and 67-99 years after the introduction of DFTD.

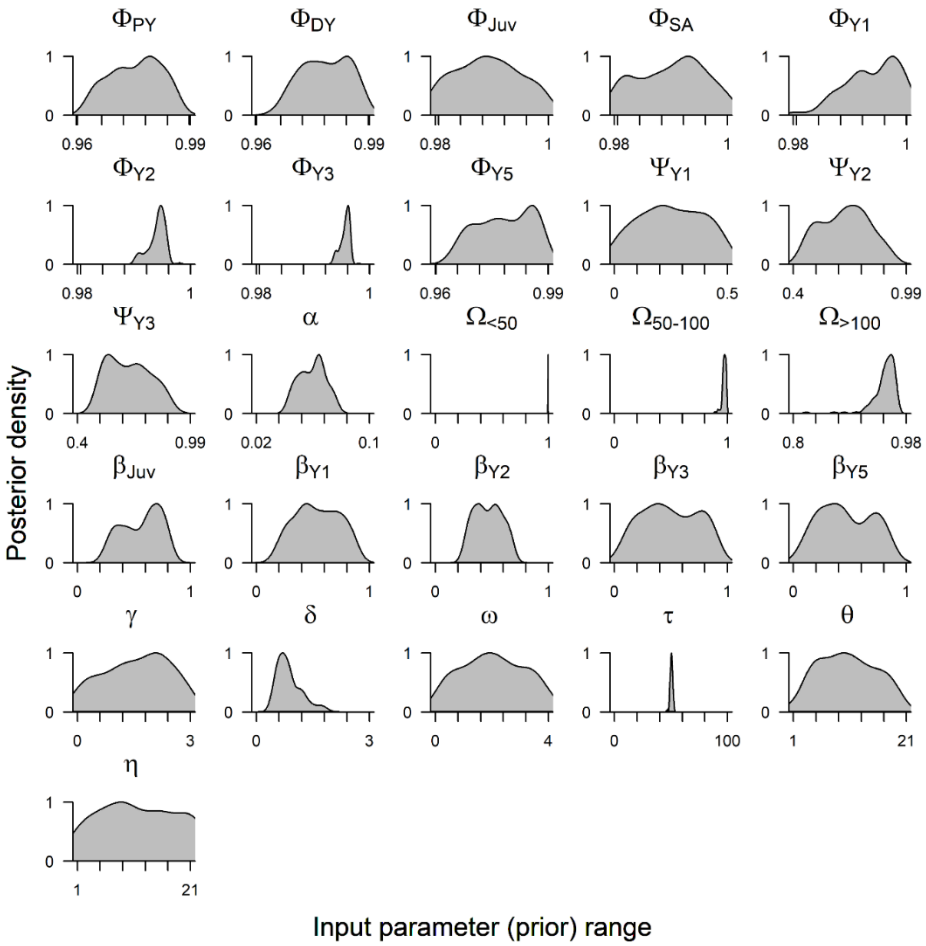

Figure S5. Posterior density distributions of parameters sampled in the individual-based model to simulate coupled demographic and epidemiological dynamics of Tasmanian devils and devil facial tumour disease (DFTD). Posterior distributions are based on selected scenarios that best matched population-level trajectories of a devil population studied over ten years. Parameter values were adjusted according to the neural network regression-based Approximate Bayesian Computation approach. The displayed ranges on the x-axes represent the prior range for each parameter (priors were drawn from uniform distributions). Symbology is described in Table S1 numbers of posterior mode and 95% credible intervals are given in Table S2. Note that the posterior distribution for $\Omega_{<50}$ is poorly visible because all values are close to 1.

Table S2. Posterior mode and 95% credible intervals (95% CI) of parameters sampled in the individual-based model and selected with Approximate Bayesian Computation. Mode and 95% CIs are presented for the adjusted parameter values (corresponding to the posterior distributions in Figure S5) and unadjusted parameter values. See Table S2 for parameter description and the full prior ranges sampled in the simulations.

Parameter	Adjusted parameters		Unadjusted parameters	
	Mode	95% CI	Mode	95% CI
Φ_{PY}	0.983	0.967 - 0.991	0.985	0.965 - 0.994
Φ_{DY}	0.988	0.971 - 0.993	0.989	0.967 - 0.995
Φ_{Juv}	0.989	0.982 - 0.997	0.989	0.983 - 0.997
Φ_{SA}	0.993	0.982 - 0.998	0.993	0.983 - 0.997
Φ_{Y1}	0.996	0.987 - 0.999	0.996	0.987 - 0.999
Φ_{Y2}	0.994	0.991 - 0.995	0.995	0.986 - 0.999
Φ_{Y3}	0.995	0.993 - 0.996	0.996	0.987 - 0.999
Φ_{Y5}	0.988	0.964 - 0.993	0.989	0.962 - 0.995
Ψ_{Y1}	0.214	0.011 - 0.491	0.225	0.019 - 0.495
Ψ_{Y2}	0.709	0.447 - 0.877	0.702	0.402 - 0.954
Ψ_{Y3}	0.561	0.493 - 0.856	0.51	0.407 - 0.95
α	0.064	0.043 - 0.075	0.075	0.024 - 0.099
$\Omega_{<50}$	0.997	0.994 – 1	0.999	0.991 – 1
$\Omega_{50\text{-}100}$	0.975	0.914 – 1	0.987	0.915 – 1
$\Omega_{>100}$	0.958	0.917 - 0.973	0.964	0.881 - 0.992
β_{Juv}	0.70	0.29 - 0.81	0.793	0.09 - 0.997
β_{Y1}	0.44	0.21 - 0.87	0.397	0.074 - 0.97
β_{Y2}	0.38	0.27 - 0.69	0.63	0.024 - 0.97
β_{Y3}	0.39	0.11 - 0.89	0.45	0.068 - 0.99
β_{Y5}	0.37	0.11 - 0.85	0.35	0.04 - 1
γ	2.08	0.006 – 3.0	2.34	0.31 - 2.94
δ	0.69	0.31 - 1.54	0.513	0 - 2.09
ω	1.93	0.13 - 3.91	1.96	0.081 - 3.83
τ	50.5	48.5 - 52.6	58.7	22.9 - 94.3*
θ	9.9	2.4 - 19.3	9.4	1.2 - 20.2
η	8.7	0.8 – 21	9.0	2.6 - 21

References

Grimm, V., U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, A. Huth, J. U. Jepsen, C. Jorgensen, W. M. Mooij, B. Mueller, G. Pe'er, C. Piou, S. F. Railsback, A. M. Robbins, M. M. Robbins, E. Rossmanith, N. Rueger, E. Strand, S. Souissi, R. A. Stillman, R. Vabo, U. Visser, and D. L. DeAngelis. 2006. A standard protocol for describing individual-based and agent-based models. Ecological Modelling **198**:115-126.

Hamede, R. K., J. Bashford, H. McCallum, and M. Jones. 2009. Contact networks in a wild Tasmanian devil (Sarcophilus harrisii) population: using social network analysis to reveal seasonal variability in social behaviour and its implications for transmission of devil facial tumour disease. Ecology Letters **12**:1147-1157.

Hamede, R. K., A.-M. Pearse, K. Swift, L. A. Barmuta, E. P. Murchison, and M. E. Jones. 2015. Transmissible cancer in Tasmanian devils: localized lineage replacement and host population response. Proceedings of the Royal Society of London B: Biological Sciences **282**:20151468.

McCallum, H., M. Jones, C. Hawkins, R. Hamede, S. Lachish, D. L. Sinn, N. Beeton, and B. Lazenby. 2009. Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-induced extinction. Ecology **90**:3379-3392.

Pye, R., R. Hamede, H. V. Siddle, A. Caldwell, G. W. Knowles, K. Swift, A. Kreiss, M. E. Jones, A. B. Lyons, and G. M. Woods. 2016. Demonstration of immune responses against devil facial tumour disease in wild Tasmanian devils. Biology letters **12**.

R Development Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Stein, M. 1981. Large sample properties of simulations using latin hypercube sampling. Technometrics **29**:143–151.

Wells, K., R. Hamede, D. H. Kerlin, A. Storfer, P. A. Hohenlohe, M. E. Jones, and H. I. McCallum. 2017. Infection of the fittest: devil facial tumour disease has greatest effect on individuals with highest reproductive output. Ecology Letters **20**:770–778.