Human Autonomy Teaming Architecture (Philosophy)

Jay Shively

NASA-Ames

Outline

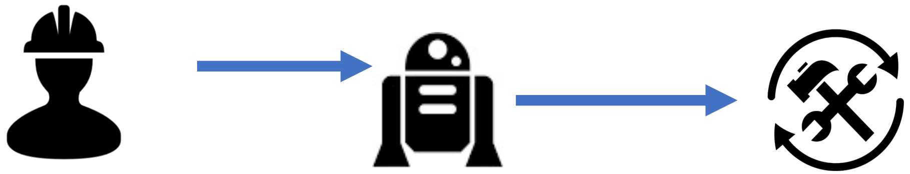
- HAT
  - What is it?
  - Why?
  - How?
- Principles
  - Bi-directional communication
  - Transparency
  - Operator directed interface
  - Shared situation awareness
  - Meaningful Human Control (MCH)
- Tools
  - Playbook
  - Working Agreements

Human & System

Human interacts directly with the work system.

Human & AI & System

AI is between the human and the work system (e.g., flight management system).

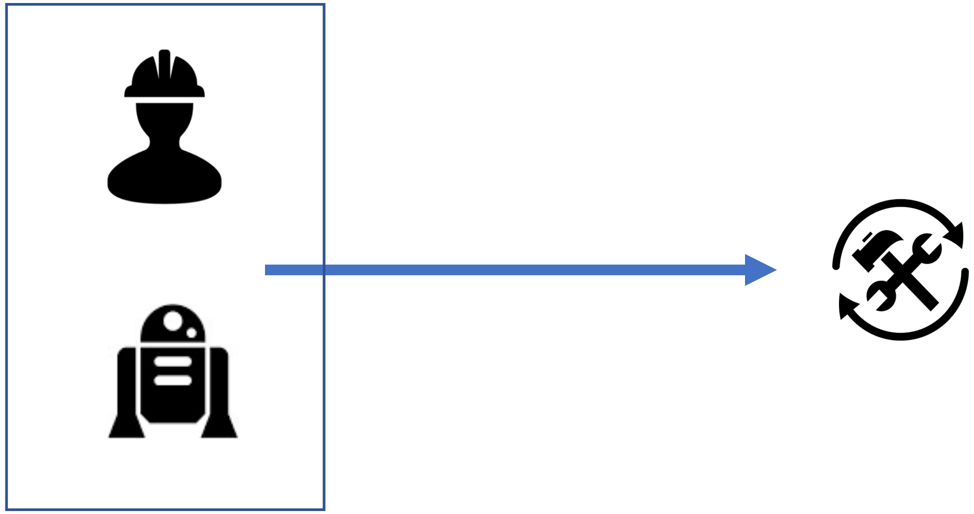


Problems with serial arrangement

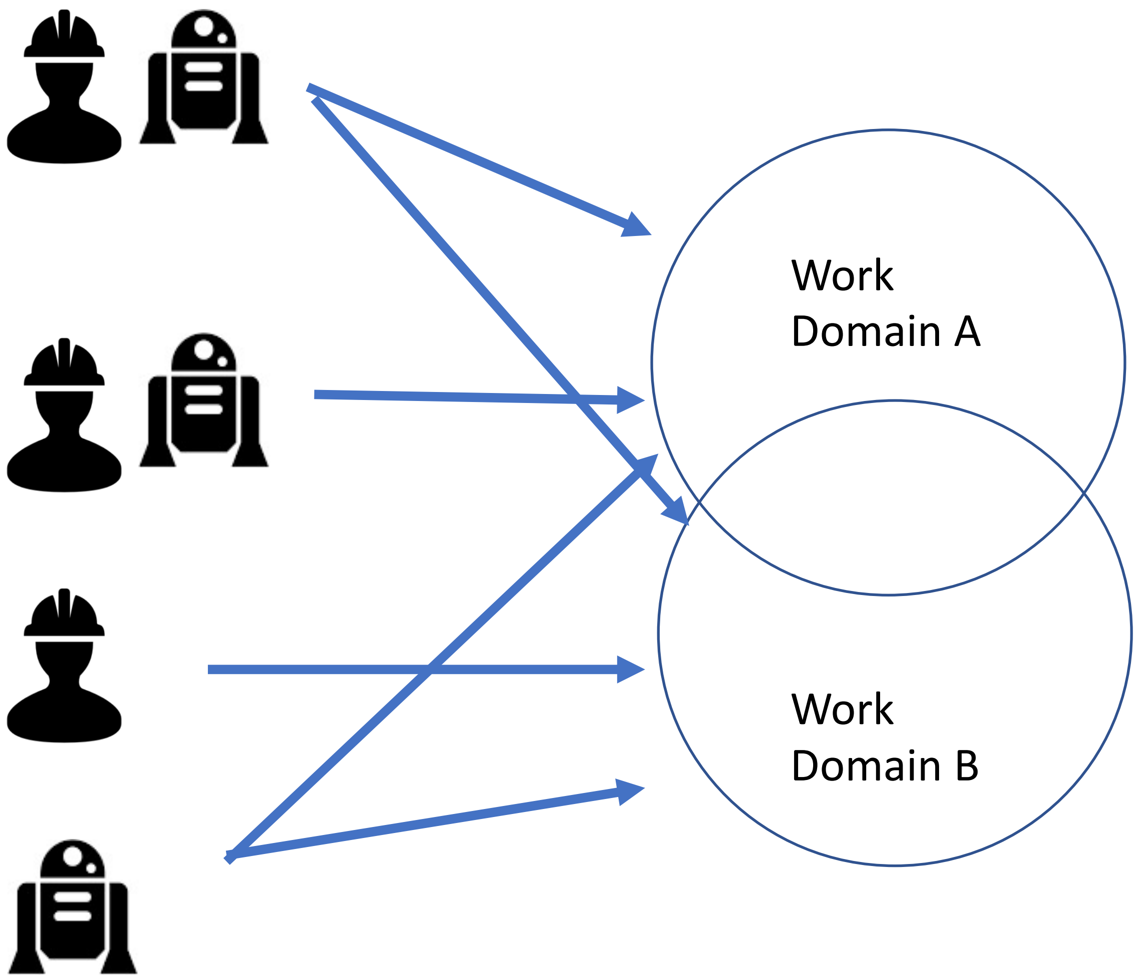
- Brittle
  - Automation often operates well for a range of situations but requires human intervention to handle boundary conditions (Woods & Cook, 2006)
- Opaque
  - Automation interfaces often do not facilitate understanding or tracking of the system (Lyons, 2013)
- Miscalibrated Trust
  - Disuse and misuse of automation have led to real-world mishaps and tragedies (Lee & See, 2004; Lyons & Stokes, 2012)
- Out-of-the-Loop Loss of Situation Awareness
  - Trade-off: automation helps manual performance and workload but recovering from automation failure is often worse (Endsley, 2016; Onnasch, Wickens, Li, Manzey, 2014)

Human AI Teaming

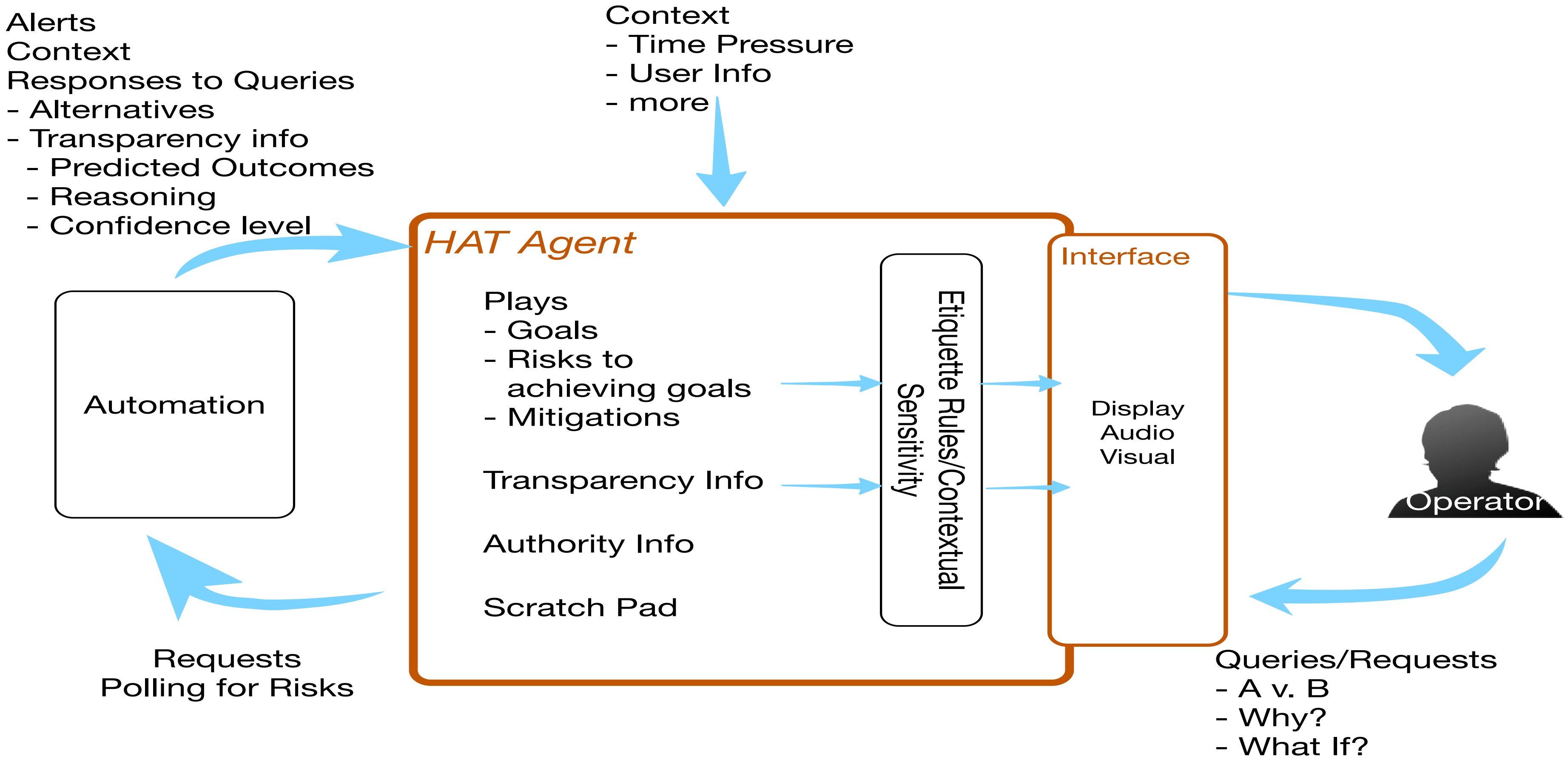
Human and AI “team” to interact with the work system.



Real World Complexity



HAT Conceptualization



Principles (Not Exhaustive) (Pros and Cons)

- Bi-directional Communication
- Transparency
- Shared situation awareness (mental model)
- Pilot directed interface
- Meaningful Human Control

Bi-Directional Communication

- Humans and AI MUST have a channel to communicate
- Human can ask about reasoning:

Why did you choose A?

What about B?

- About confidence:

How sure are you of A? How close was B?

- Human can input information that the AI doesn't have

Alternative C doesn't have good emergency medical response

Bi-Directional Communication

- Human must be able to bounce ideas off the AI (and vice versa) like a TEAMMATE

- These capabilities do levy requirements on the AI

  - Common/understandable comm mode
  - Natural language interface (desired but not required)
  - Explainability
  - Self-awareness? Level of confidence

Transparency

- Human MUST be able to understand the reasoning for AI recommendations/decisions

- Resulting in calibrated trust of the AI

- Resulting in appropriate utilization

Transparency

- Issues
  - Granularity
  - Time pressure

Shared Situation Awareness

- Need Shared
  - Mental model of system operation
  - State of the system
  - State of the environment
- How do you ensure?
- How do you transfer info, if not present?
- How do you measure?

Operator Directed Interface

- No set “roles and responsibilities”
  - Default assignments for nominal operations
- Fluid assignment set by operator actions
  - Take control of stick and fly aircraft
- Requires some degree of ”Intent Inferencing” – can be problematic
- Can be difficult to implement in highly automated systems

Meaningful Human Control

- Humans must have the ability to exert “meaningful” control
    - Not just hit the red button
- Take advantage of human decision making and analytical skills
- Ethical, moral, humanistic reasons

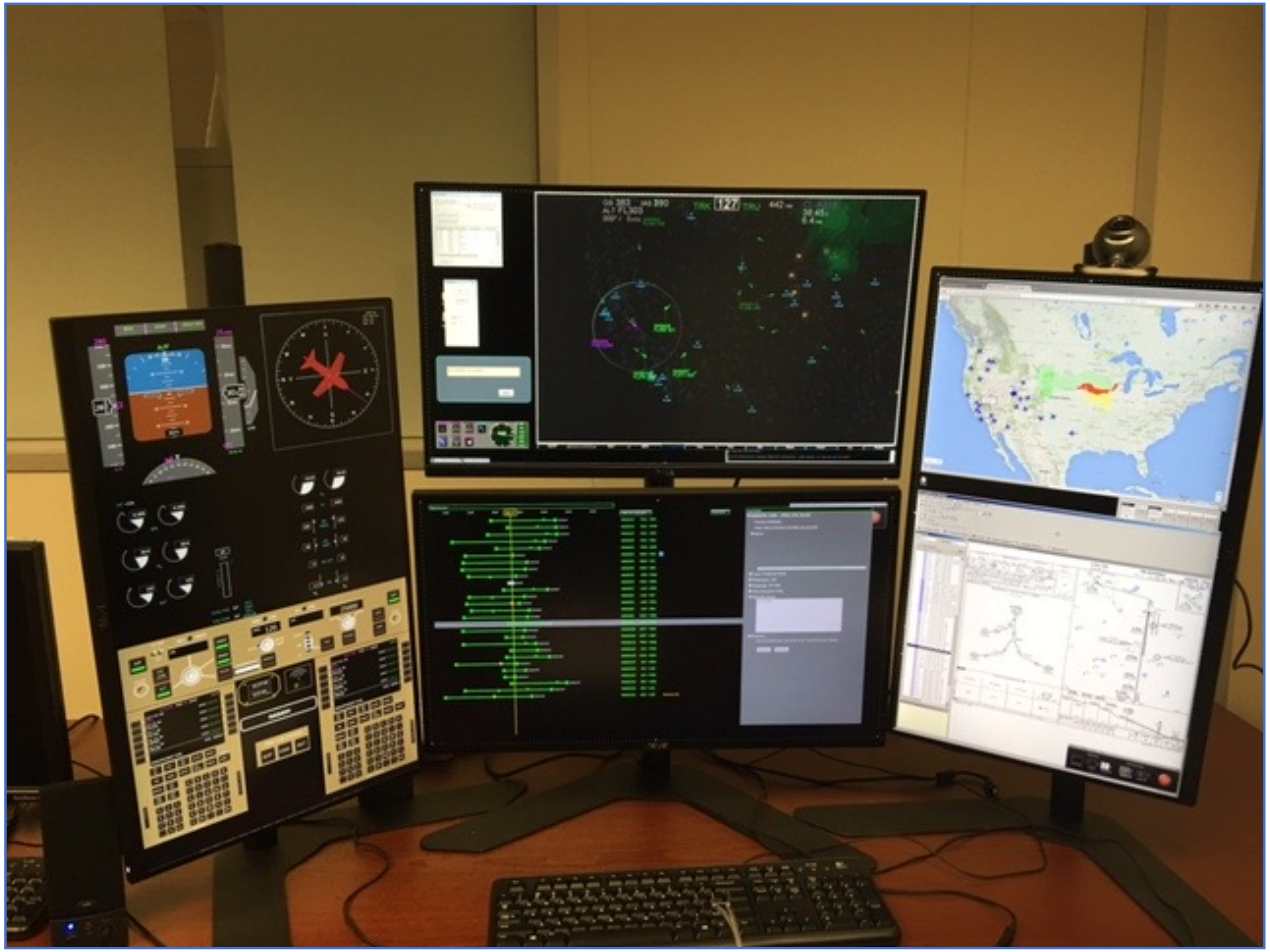
MHC

- Issues
  - How do you define
  - How do you measure
  - Time delay
  - Human limitations

Applications

- Single Pilot Operations
- m:N Vehicle Control

Reduced Crew Operations



ELP and ACFP

Research prototype software, Intelligent Systems Division, PI: D. Smith

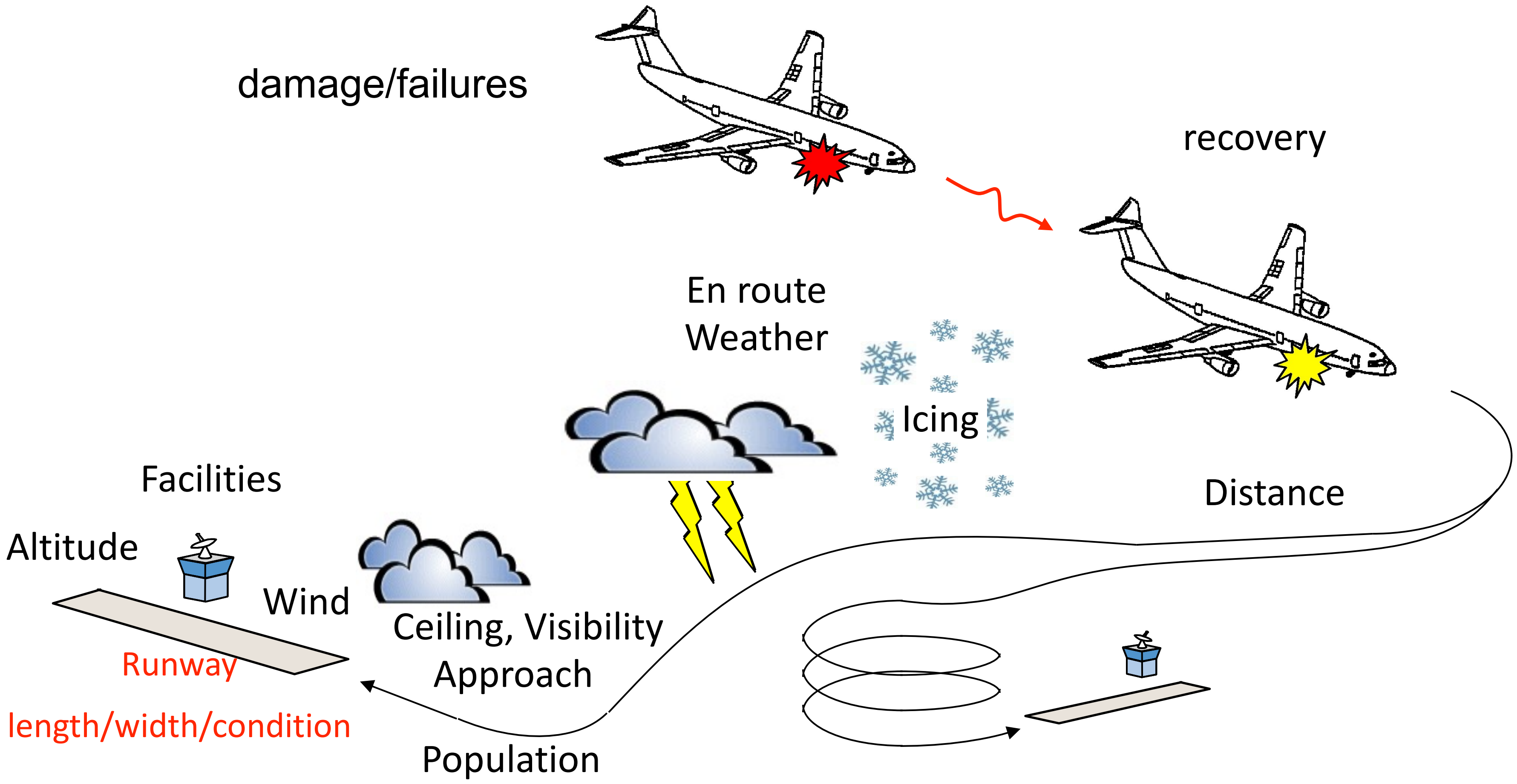
ELP – Emergency Landing Planner (2007-2012)

- Cockpit decision aid
- Route planning for (serious) emergencies
  - control system failures
  - physical damage
  - fires
- Time & Safety were dominant considerations

ACFP – Autonomous Constrained Flight Planer (2013-2017)

- Ground station decision aid
- Diversion selection, route planning, route evaluation
  - weather diversion
  - medical emergencies
  - less critical system failures

ELP Objective



Find the best landing sites and routes for the aircraft

ELP Approach

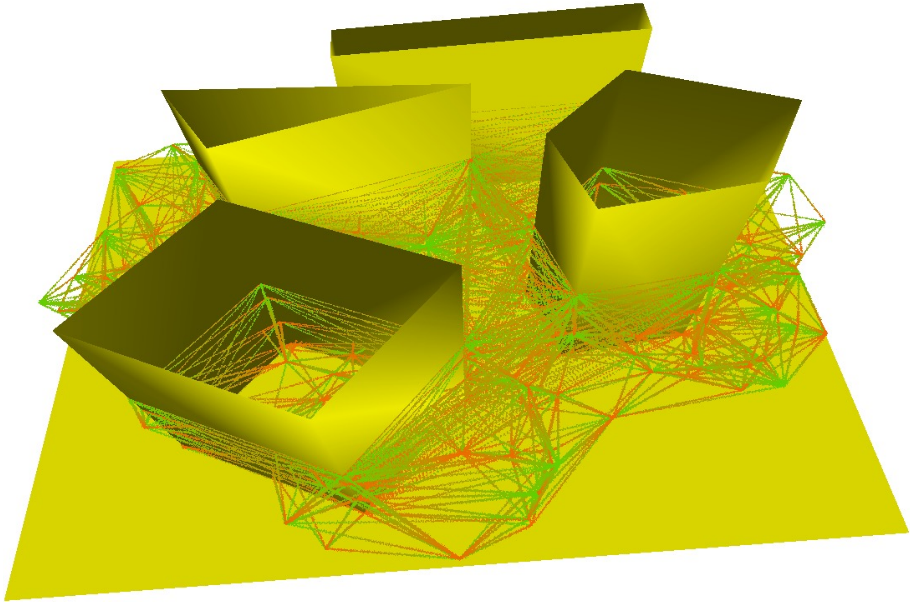
Consider all runways within range (150 miles)

Construct “obstacles” for weather & terrain

Search for paths to each runway

Evaluate risk of each path

Present ordered list



< 10 seconds

ELP’s Risk Model

Enroute path
- Distance/time
- Weather

Approach path
- Ceiling & Visibility
- Approach minimums
- Population density

Runway
- Length
- Width
- Surface condition
- Relative wind

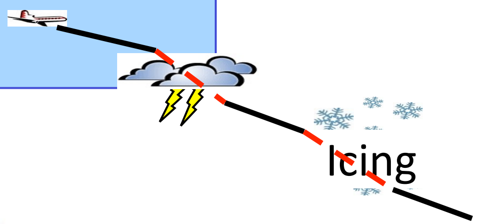
Airport
- Density altitude
- Tower
- Weather reporting
- Emergency facilities

P_{stable} ≡ probability of success / nm in stable flight

P_{wx} ≡ probability of success / nm in light weather

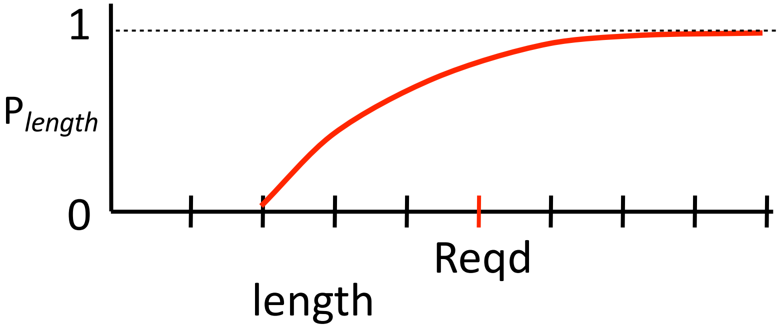
$P_{leg} \equiv (P_{stable} * (P_{wx})^S)^D$

$P_{route} \equiv \prod P_{leg}$

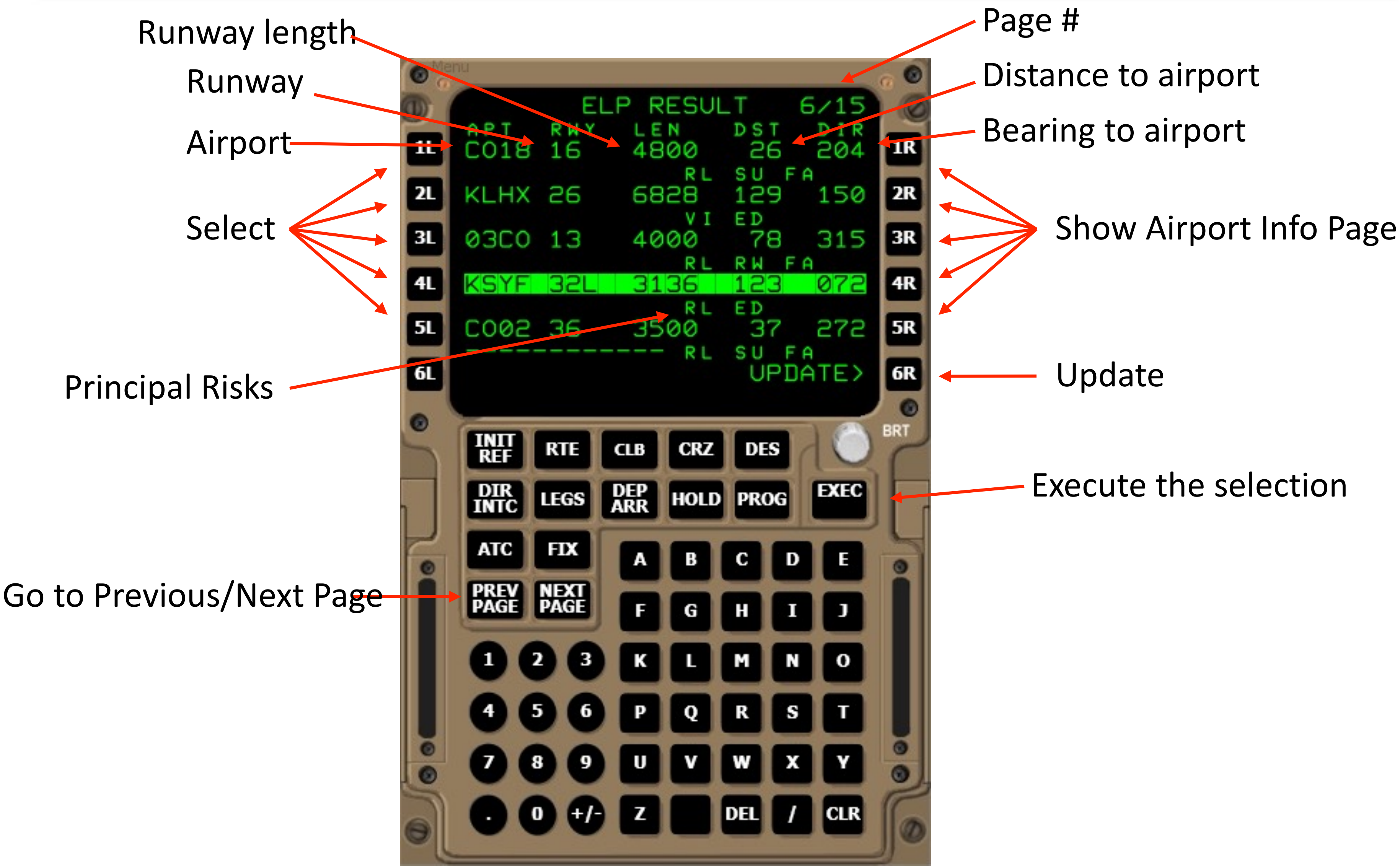


$P_{appr} \equiv P_{leg} * P_{ceil} * P_{vis}$

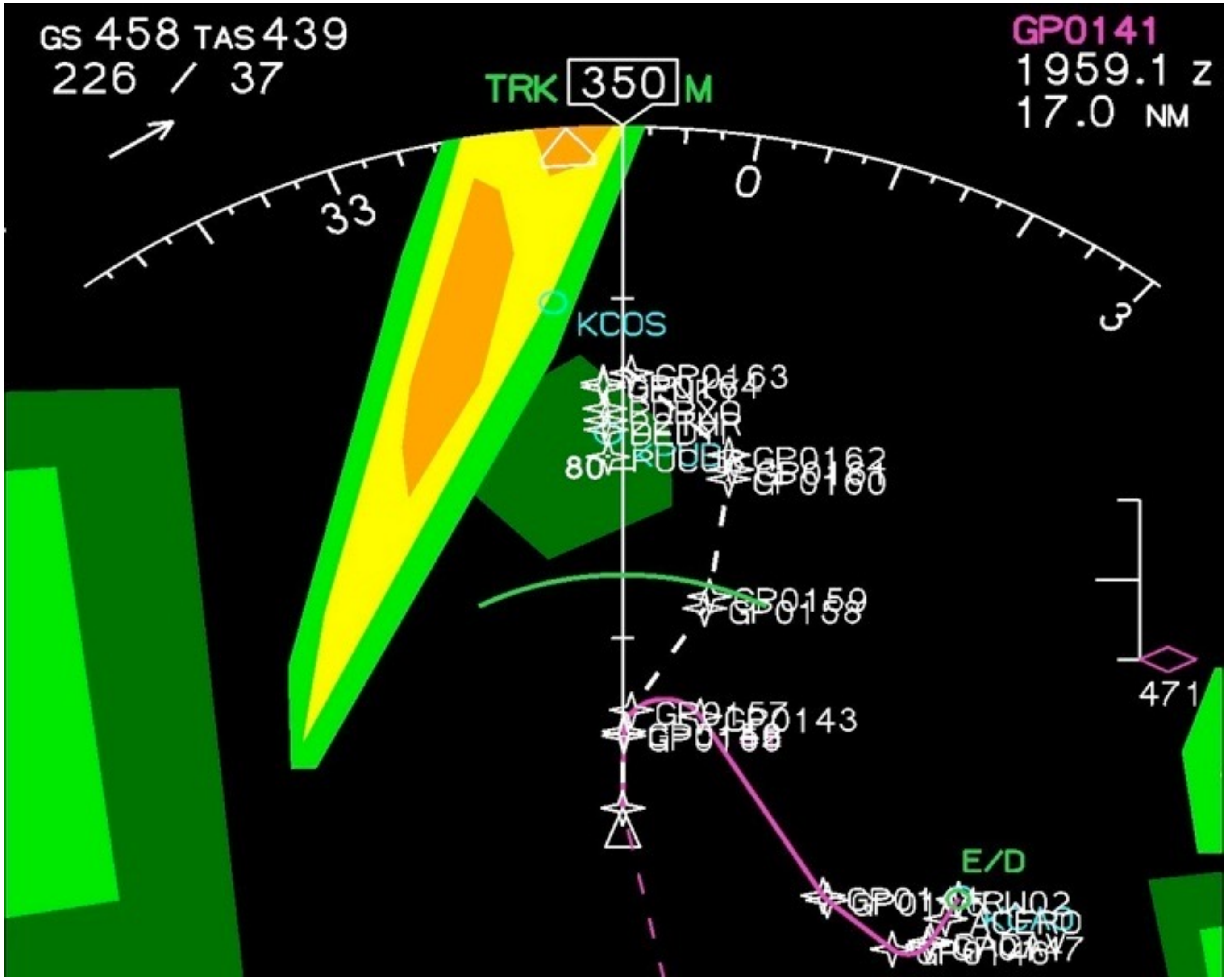
$P_{rnwy} \equiv P_{length} * P_{width} * P_{surf} * P_{speed} * P_{xwind}$



Emergency Page on the CDU



ELP Routes on the Navigation Display



ELP Experiment (2010)

Evaluation of ELP in ACFS

- 3 physical damage scenarios
- 5 pilot teams
- 16 scenarios each

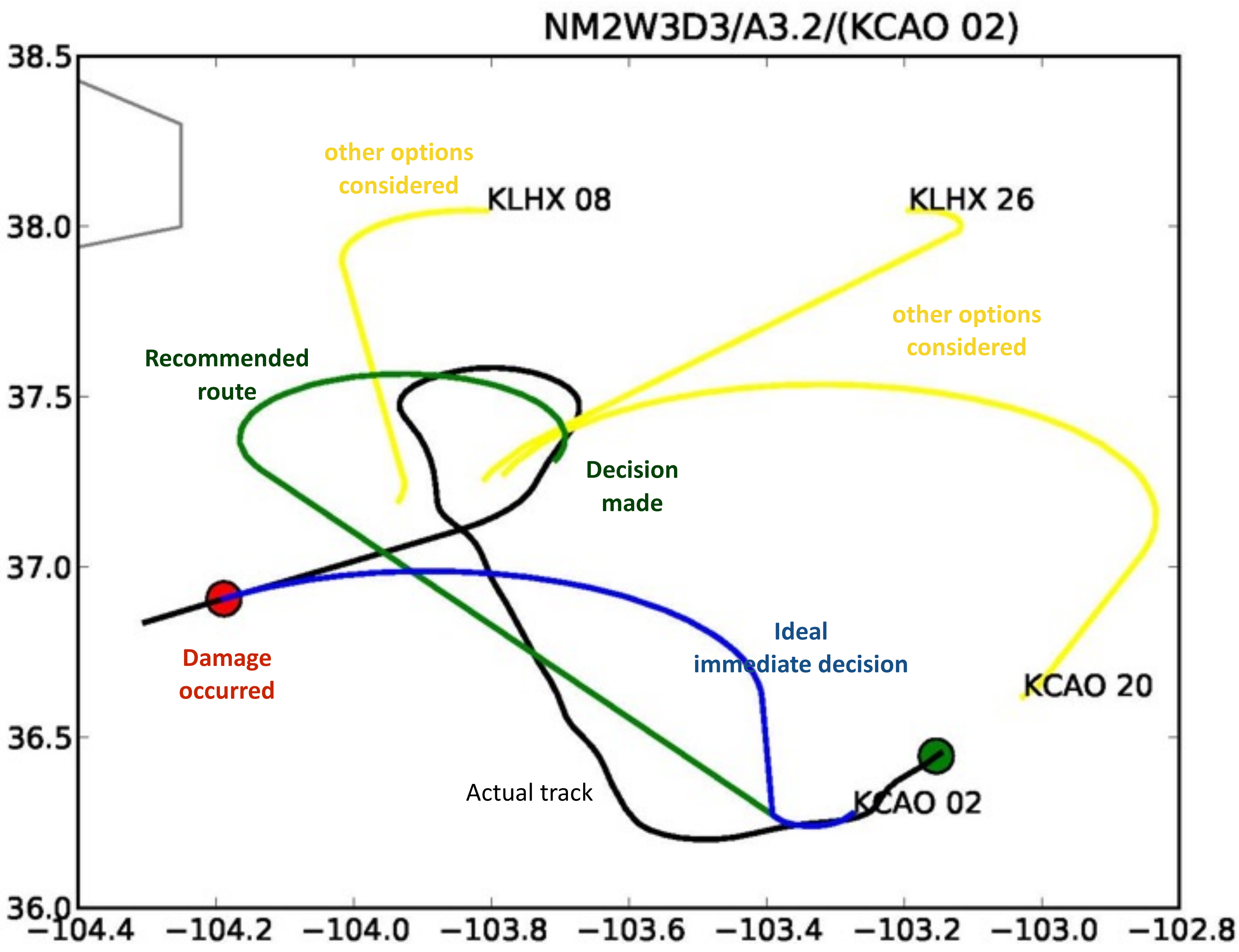
Results

- Decision **quality** somewhat better in adverse weather
- Decision **speed** much better in adverse weather
- Damage Severity not a significant factor

Pilot feedback:

“ **...** your software program alleviates the uncertainty about finding a suitable landing site and also reduces workload so the crew can concentrate on "flying" the aircraft.”

Sample Run



NM2W3D3/A3.2/(KCAO 02)
38.5
38.0
37.5
37.0
36.5
36.0
−104.4
−104.2
−104.0
−103.8
−103.6
−103.4
−103.2
−103.0
−102.8
other options considered
KLHX 08
KLHX 26
other options considered
Recommended route
Decision made
Damage occurred
Ideal immediate decision
KCAO 20
KCAO 02
Actual track

ACFP differences

Multiple aircraft

Much wider geographic area

Additional optimization criteria

- medical facilities
- maintenance facilities
- passenger facilities
- connections

Constrained requests

- runway length
- distance

Route evaluation

- current route/destination
- proposed changes

RCO Ground station

Optimization

Situations:

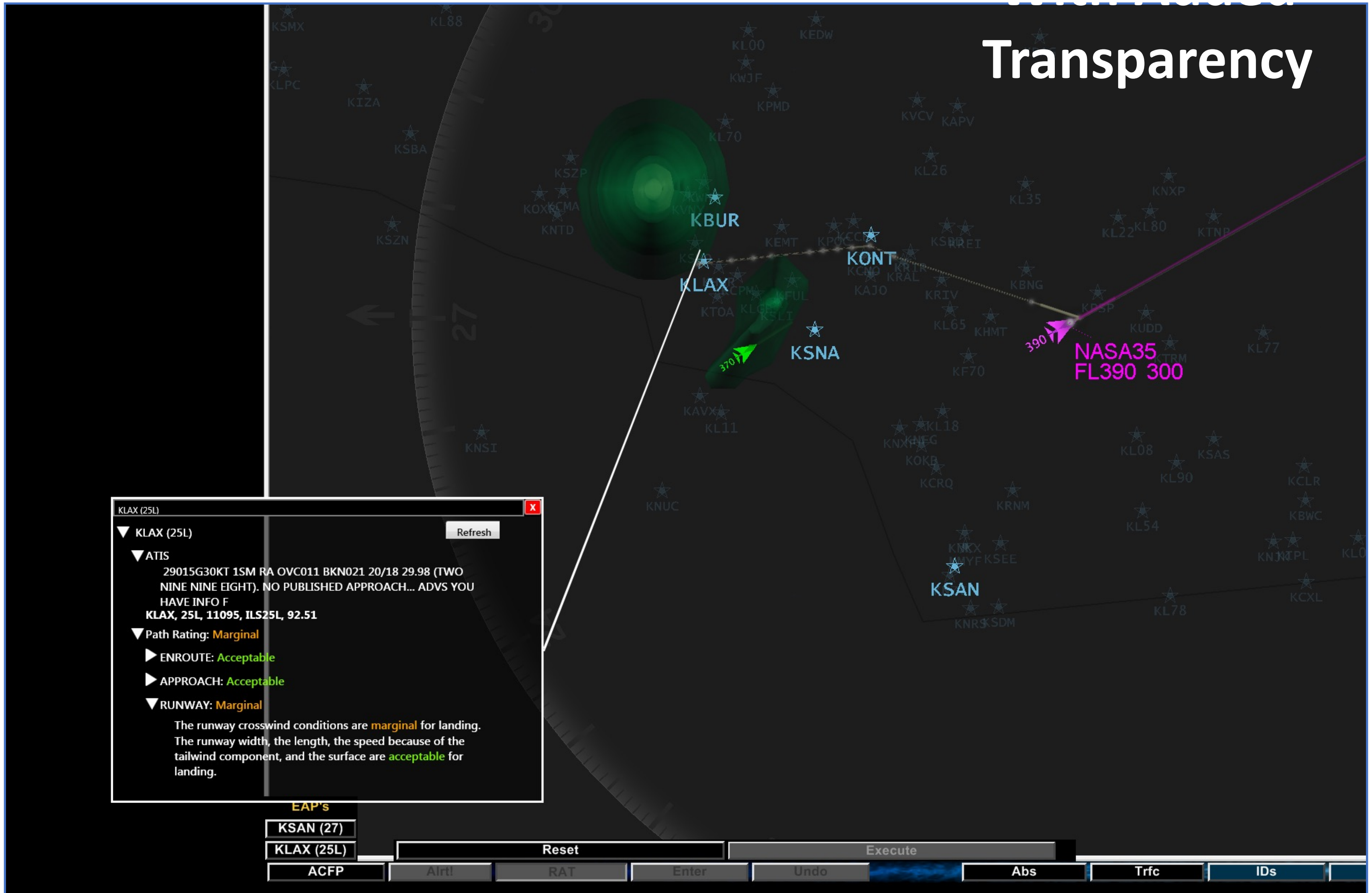
- weather reroute
- weather diversion
- systems diversion
  - anti-skid braking
  - radar altimeter
- medical emergency
  - heart attack
  - laceration
- engine loss
- depressurization
- damage
- cabin fire

Safety	Time	Medical	Conven.	Maint.
★	★		★	
★			★	
★			★	★
★			★	★
★		★	★	
★		★	★	
★	★		★	★
★	★			★
★	★			
★	★			

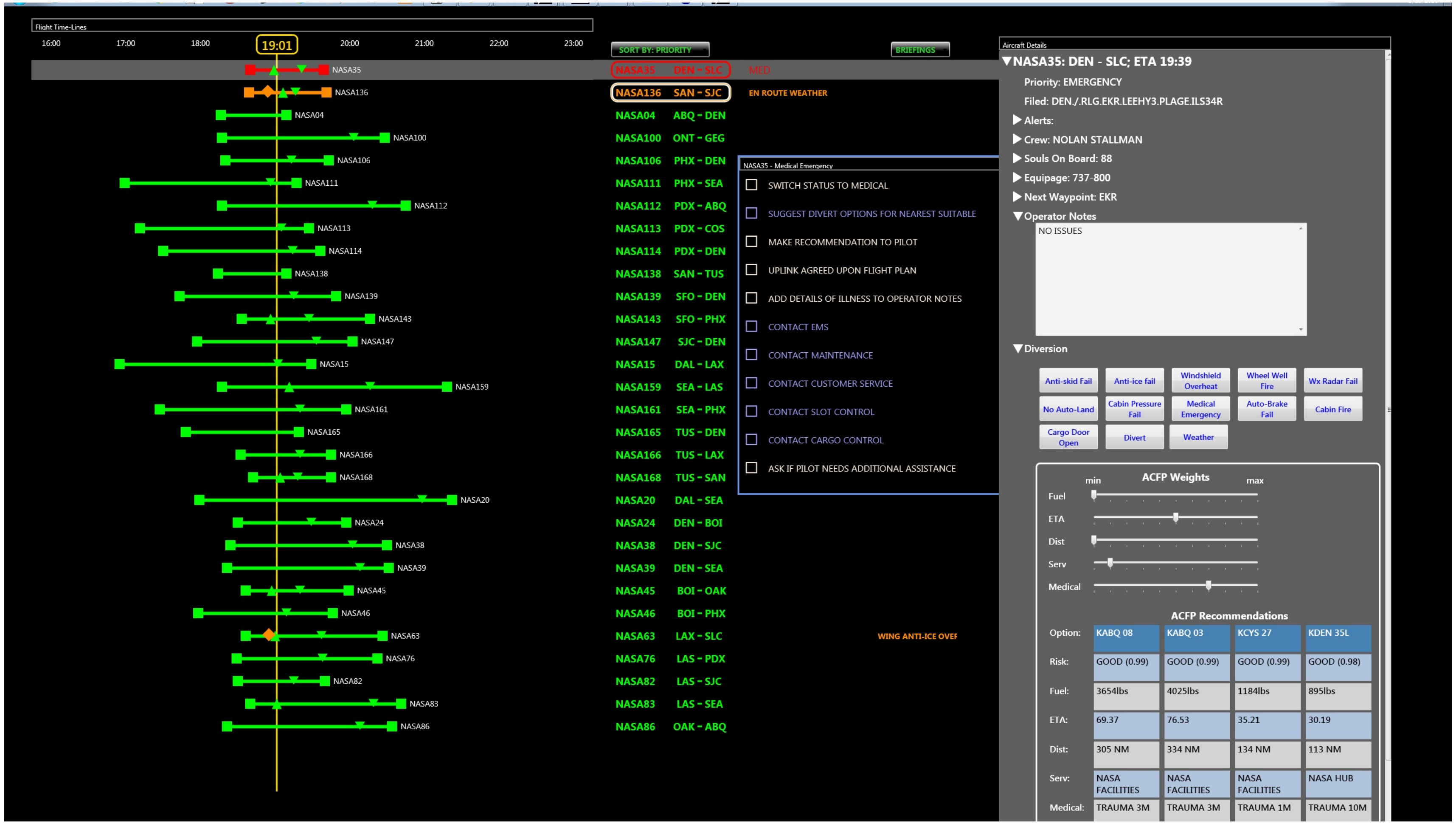
ACFP Before HAT

Recommended airports - rank ordered.

Adding HAT Principles to the Ground Station

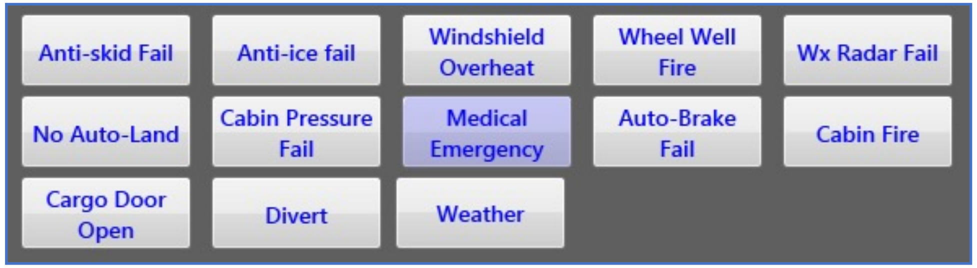


Adding HAT Principles to the Ground Station

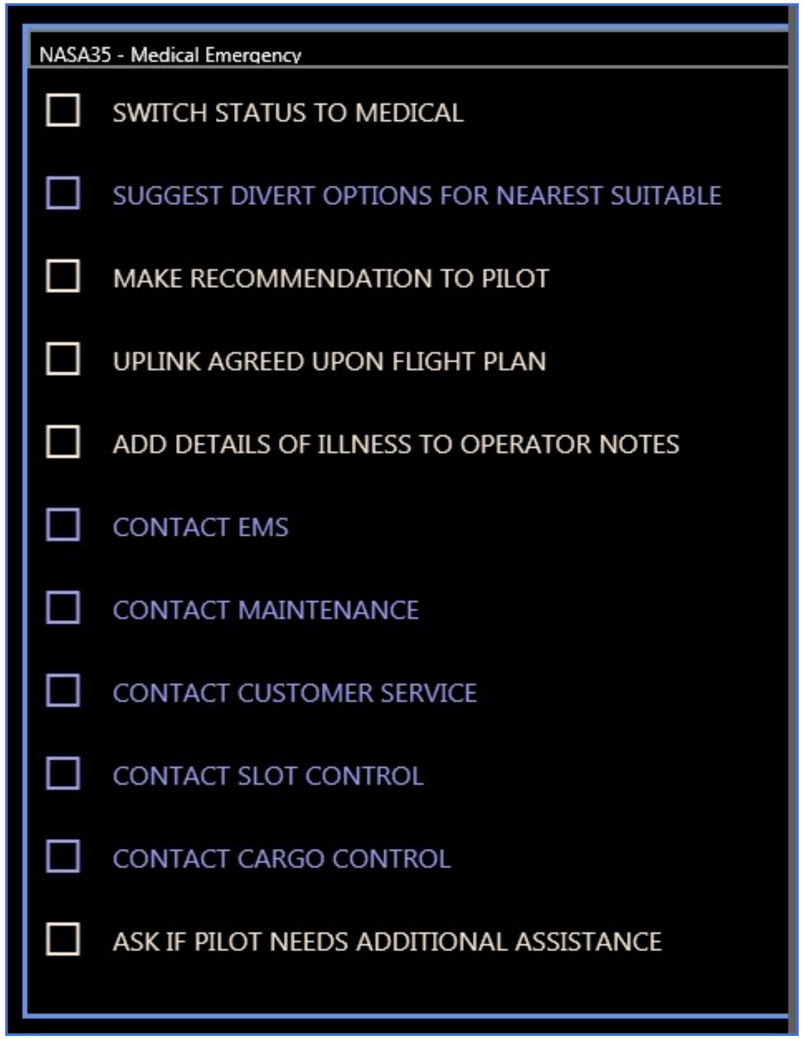


Adding HAT Principles to the Ground Station

- Human-Directed: Operator calls “Plays” to delegate authority to automation



A play encapsulates a plan for achieving a goal.
It includes roles and responsibilities
    what is the automation going to do
    what is the operator going to do

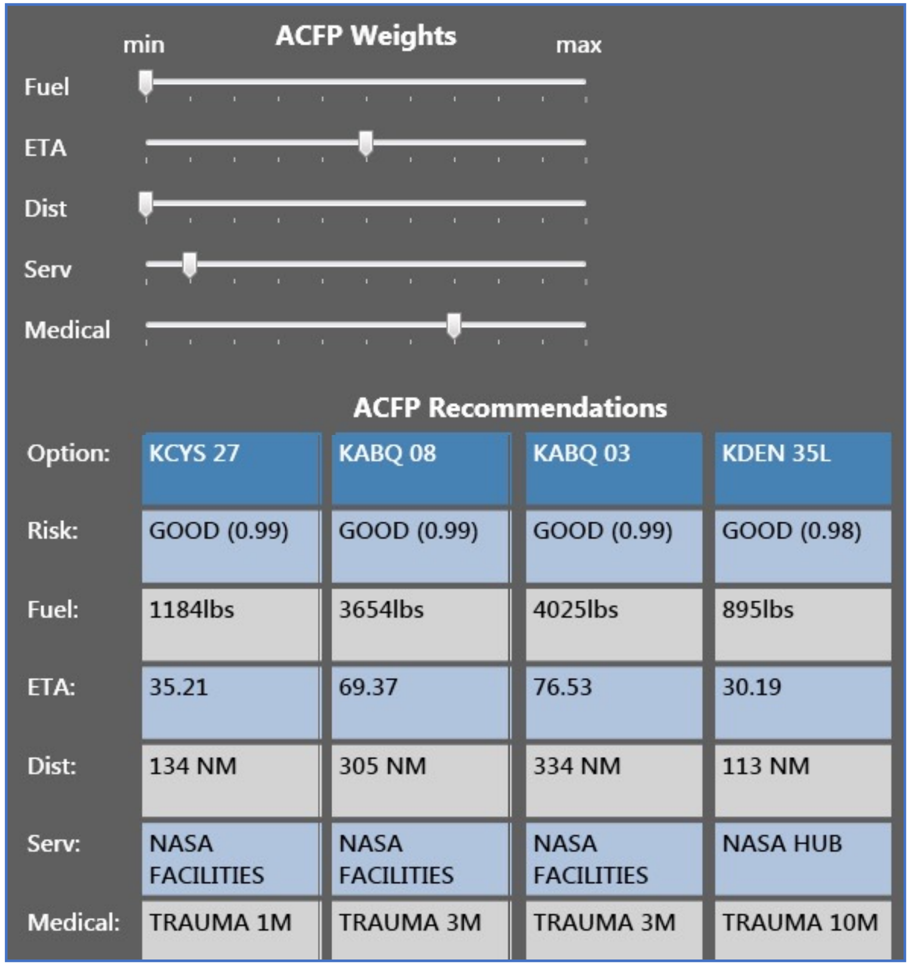


Adding HAT Principles to the Ground Station

- Transparency: Divert reasoning and factor weights are displayed.

- Negotiation/Dialog: Operators can change factor weights to match their priorities.

- Shared Language/Communication: Numeric output from ACFP was found to be misleading by pilots. Display now uses English categorical descriptions.



HAT Simulation: Tasks

- Participants, with the help of automation, monitored 30 aircraft
  - Alerted pilots when
    - Aircraft was off path or pilot failed to comply with clearances
    - Significant weather events affect aircraft trajectory
    - Pilot failed to act on EICAS alerts
  - Rerouted aircraft when
    - Weather impacted the route
    - System failures or medical events force diversions

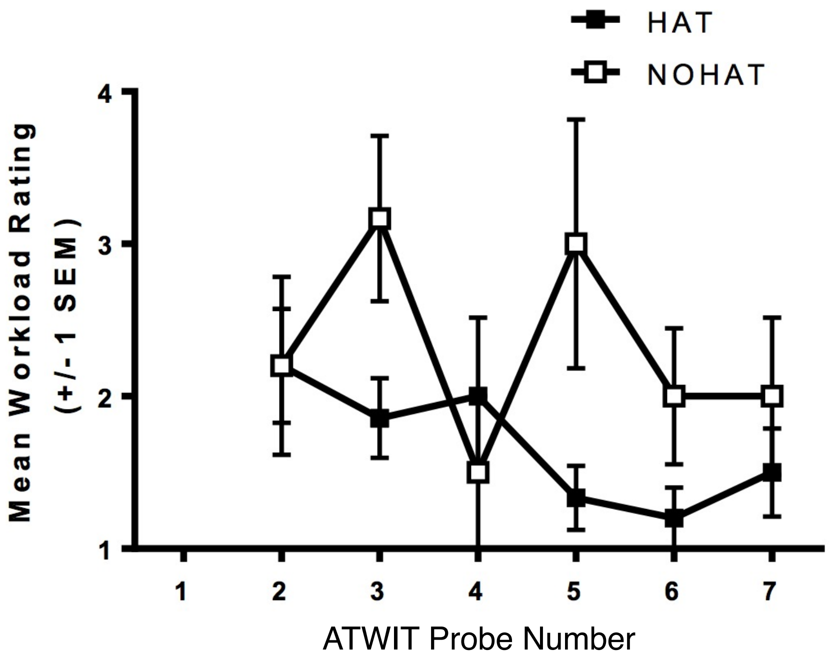
- Ran with HAT tools and without HAT tools

HAT Simulation: Results

- Participants preferred the HAT condition overall (rated 8.5 out of 9).
- HAT displays and automation preferred for keeping up with operationally important issues (rated 8.67 out of 9)
- HAT displays and automation provided enough situational awareness to complete the task (rated 8.67 out of 9)
- HAT displays and automation reduced the workload relative to no HAT (rated 8.33 out of 9)

HAT Simulation: Results

- HAT workload reduction was marginally significant (HAT mean 1.7; No HAT mean 2.3, p = .07)



HAT Simulation: Debrief

- Transparency
  - “This [the recommendations table] is wonderful…. You would not find a dispatcher who would just be comfortable with making a decision without knowing why.”

- Negotiation
  - “The sliders was [sic] awesome, especially because you can customize the route…. I am able to see what the difference was between my decision and [the computer’s decision].”

- Human-Directed Plays/Shared Plans
  - “Sometimes [without HAT] I even took my own decisions and forgot to look at the [paper checklist] because I was very busy, but that didn’t happen when I had the HAT.”

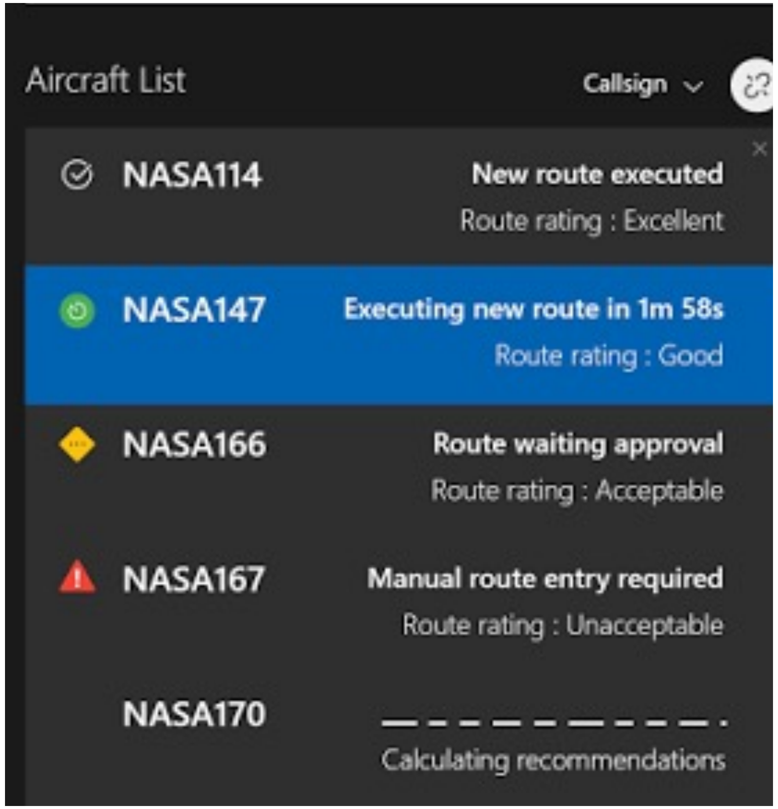
HAT Simulation: Summary

- Participants liked where we were headed with the HAT concept
  - Increased Situation Awareness
  - Reduced Workload
- Things we didn’t get quite right
  - Annunciations: People liked them but thought there were to many
  - Voice Control: Did not work well. Need a more complete grammar, better recognition
  - Participants didn’t always understand what the goal of a play was
- Things we didn’t get to
  - Airlines hate diverts. We need to put in support to help avoid them
  - Plays need more structure (branching logic)
  - Roles and responsibilities need to be more flexible
  - Limited ability to suggest alternatives

ALTA Action Phase

Based on working agreements, the Agent will do one of the following:

- [Auto] autonomously executes and informs operator
- [Veto] presents a solution which will be autonomously executed unless the operator intervenes
- [Select] presents multiple options for operator selection



Working Agreements:
A Path to Full Autonomy

- Codified levels of automation

- Working agreements allow the task structure to remain the same while the involvement of human operators decreases due to improvements in the automation and increases in trust

M:N Aircraft Operations

Summary

- AI as a TEAMMATE
- Basic Principles
- Real World Application